

Lattices of Graphical Gaussian Models with Symmetries

Helene Gehrmann*

Department of Statistics, University of Oxford

January 19, 2013

Abstract

In order to make graphical Gaussian models a viable modelling tool when the number of variables outgrows the number of observations, Højsgaard and Lauritzen (2008) introduced model classes which place equality restrictions on concentrations or partial correlations. The models can be represented by vertex and edge coloured graphs. The need for model selection methods makes it imperative to understand the structure of model classes. We identify four model classes that form complete lattices of models with respect to model inclusion, which qualifies them for an Edwards–Havránek model selection procedure (Edwards and Havránek, 1987). Two classes turn out most suitable for a corresponding model search. We obtain an explicit search algorithm for one of them and provide a model search example for the other.

Keywords: Conditional independence; Covariance selection; Invariance; Model selection; Patterned covariance matrices; Permutation symmetry

1 Introduction

Graphical models are probabilistic models which use graphs to represent dependencies between random variables. This article is concerned with models represented by undirected graphs, in which each variable corresponds to a vertex and a pair of vertices is joined by an edge unless the corresponding variables are conditionally independent given the remaining variables. In addition to providing a concise form of visualisation of the conditional independence structure of a model, the graphical representation can be exploited to make statistical inference computations more efficient (Lauritzen, 1996).

Motivated by the growing need for parsimonious models in modern day applications, in particular when the number of variables outgrows the number of observations, in recent years graphical models with additional equality constraints on model parameters are becoming of increasing interest, in discrete models (Gottard et al., 2010; Ramírez-Aldana, 2010) as well as in multivariate Gaussian models, which are the central object of interest in this article. First studies (Højsgaard and Lauritzen, 2008; Uhler, 2010) show that equality constraints reduce the minimal number of observations required to ensure estimability of the model parameters with probability one, which makes graphical Gaussian models with equality constraints a promising model class.

*Address for correspondence. Helene Gehrmann, Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, United Kingdom, gehrmann@stats.ox.ac.uk.

Symmetry constraints, induced by distribution invariance under a permutation group applied to the variable labels, are a special instance of equality constraints and have a long history for the Gaussian distribution before the advent of graphical models (Wilks, 1946; Votaw, 1948; Olkin and Press, 1969; Olkin, 1972; Andersson, 1975; Jensen, 1988). First studies of models combining symmetry constraints with conditional independence relations are given in Hylleberg et al. (1993); Andersen et al. (1995); Madsen (2000). The models we study have been introduced in Højsgaard and Lauritzen (2008) and contain the models in Hylleberg et al. (1993) as a special case. The types of restrictions are: equality between specified elements of the concentration matrix (RCON) and equality between specified partial correlations (RCOR). The models can be represented by vertex and edge coloured graphs, where parameters associated with equally coloured vertices or edges are restricted to being identical.

In order for RCON and RCOR models to become widely applicable in practice model selection methods need to be developed, which motivates the study of model structures. This is the main objective of this article. As both model types RCON and RCOR models form complete lattices, both qualify for the Edwards–Havránek model selection procedure for lattices (Edwards and Havránek, 1987). However due to the large number of models it is more feasible to search, at least initially, in suitable subsets of the model space. Particularly favourable are subsets of models in which equality constraints are placed in a pattern which makes them more readily interpretable.

Four model classes with desirable statistical properties which express themselves in regularity of graph colouring have been previously identified in the literature: The most restrictive is given by *graphical symmetry models* studied in Hylleberg et al. (1993), also appearing in Højsgaard and Lauritzen (2008) under the name *RCOP models*. The corresponding graph colourings are given by vertex and edge orbits of a permutation group acting on the variable labels and we therefore term them *permutation-generated*. Colourings representing models which place the same equality constraints on the concentrations and partial correlations were termed *edge regular* in Gehrmann and Lauritzen (2011). Two further model types, ensuring estimability of a non-zero model mean subject to equality constraints, were introduced in Gehrmann and Lauritzen (2011), the colourings representing them termed *vertex regular* and *regular* respectively.

The main results presented in this article are that each of the model classes forms a complete lattice of models and the identification of their meet and join operations. The former is established by showing that each model class is stable under model intersection, which gives the shared meet operation, and by demonstrating that whenever a model does not fall inside a given class there is a unique smallest larger model, or supremum, which does, giving the distinct join operations. The found lattice structure qualifies each model class for an Edwards–Havránek model search, giving a first model selection procedure for RCON and RCOR models.

We focus on models represented by edge regular and permutation-generated colourings as their structure is generally more tractable and their constraints more readily interpretable. We present an Edwards–Havránek model selection algorithm for models with edge regular colourings and illustrate it by means of an example with five variables, with a very encouraging performance. We further provide an example of a model search within models with permutation-generated colourings with four variables, commonly known as *Fret’s heads* (Frets, 1921; Mardia et al., 1979).

2 Preliminaries and Notation

2.1 Notation

Let $G = (V, E)$ be an undirected uncoloured graph with vertex set V and edge set E . For a $|V| \times |V|$ matrix $A = (a_{\alpha\beta})$, $A(G)$ shall denote the matrix defined by $A(G)_{\alpha\beta} = 0$ whenever there is no edge between α and β in G for $\alpha \neq \beta$, and $A(G)_{\alpha\beta} = a_{\alpha\beta}$ otherwise. For a set of matrices M we let M^+ denote the set of positive-definite matrices inside M . \mathcal{S} shall denote the set of symmetric matrices, so that \mathcal{S}^+ denotes the set of symmetric positive definite matrices and $\mathcal{S}^+(G)$ the set of symmetric positive definite matrices indexed by V whose $\alpha\beta$ -entry is zero for $\alpha\beta \notin E$ for $\alpha \neq \beta$. We indicate that a matrix is symmetric by only writing its elements on the diagonal and above. Asterisks as matrix entries indicate that the corresponding entry is unconstrained apart from any restrictions stated explicitly.

For a discrete set D we let $S(D)$ denote the set symmetric group acting on D , consisting of all permutations of the elements in D . For $F \subseteq S(D)$, $\langle F \rangle$ denotes the group generated by F , containing all permutation which can be expressed as products of elements in F and their inverses. Permutations are written in cycle notation, meaning that $\sigma = (i_1 i_2 \dots i_r)$ maps i_j to i_{j+1} for $1 \leq j < r$ and i_r to i_1 .

For a graph $G = (V, E)$, $\text{Aut}(G)$ denotes the automorphism group of G , containing all permutations in $S(V)$ which leave G invariant. For a partition P of a set S and $a, b \in S$ we write $a \equiv b (P)$ to denote that a and b lie in the same set in P . For $n \in \mathbb{N}$, we denote sets of the form $\{1, 2, \dots, n\}$ by $[n]$.

2.2 Graphical Gaussian Models

A *graphical Gaussian model* is concerned with the distribution of a multivariate random vector $Y = (Y_\alpha)_{\alpha \in V}$. Let $G = (V, E)$ be an undirected graph with vertex set V and edge set E . Then the graphical Gaussian model represented by G is given by assuming that Y follows a multivariate Gaussian $\mathcal{N}_{|V|}(\mu, \Sigma)$ distribution with concentration matrix $K = \Sigma^{-1} \in \mathcal{S}^+(G)$.

The entries in $K = (k_{\alpha\beta})_{\alpha, \beta \in V}$ have a simple interpretation. The diagonal elements $k_{\alpha\alpha}$ are reciprocals of the conditional variances given the remaining variables

$$k_{\alpha\alpha} = \text{Var}(Y_\alpha | Y_{V \setminus \{\alpha\}})^{-1}$$

for $\alpha \in V$. The scaled elements of the concentration matrix

$$c_{\alpha\beta} = \frac{k_{\alpha\beta}}{\sqrt{k_{\alpha\alpha}k_{\beta\beta}}} \tag{1}$$

for $\alpha, \beta \in V$ are the negative partial correlation coefficients

$$\rho_{\alpha\beta|V \setminus \{\alpha, \beta\}} = \frac{\text{Cov}(Y_\alpha, Y_\beta | Y_{V \setminus \{\alpha, \beta\}})}{\text{Var}(Y_\alpha | Y_{V \setminus \{\alpha\}})^{1/2} \text{Var}(Y_\beta | Y_{V \setminus \{\beta\}})^{1/2}} = -c_{\alpha\beta} \tag{2}$$

for $\alpha, \beta \in V$. It follows that

$$\alpha\beta \notin E \iff k_{\alpha\beta} = 0 \iff Y_\alpha \perp\!\!\!\perp Y_\beta \mid Y_{V \setminus \{\alpha, \beta\}} \tag{3}$$

see e.g. Chapter 5 in Lauritzen (1996) for further details.

2.3 Graph Colouring

For general graph terminology we refer to Bollobás (1998). Following Højsgaard and Lauritzen (2008), for $G = (V, E)$ an undirected graph, a *vertex colouring* of G is a partition $\mathcal{V} = \{V_1, \dots, V_k\}$ of V , where we refer to V_1, \dots, V_k as the *vertex colour classes*. Similarly, an *edge colouring* of G is a partition $\mathcal{E} = \{E_1, \dots, E_l\}$ of E into l *edge colour classes* E_1, \dots, E_l . A colour class with a single element is *atomic* and a colour class which is not atomic is *composite*. We let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote the *coloured graph* with vertex colouring \mathcal{V} and edge colouring \mathcal{E} and let $(\mathcal{V}, \mathcal{E})$ denote its *graph colouring*. For \mathcal{V} and \mathcal{E} as above and $u \in \mathcal{V}$, we let T^u denote the $|V| \times |V|$ diagonal matrix with $T^u_{\alpha\alpha} = 1$ if and only if $\alpha \in u$ and zero otherwise. Similarly for $u \in \mathcal{E}$, we let T^u be the symmetric $|V| \times |V|$ matrix with $T^u_{\alpha\beta} = 1$ if and only if $\alpha\beta \in u$ and zero otherwise.

In our display of vertex and edge coloured graphs, we indicate the colour class of a vertex by the number of asterisks we place next to it. Similarly we indicate the colour class of an edge by dashes. Vertices and edges which are displayed in black correspond to atomic colour classes.

2.4 Lattices

A *binary relation* ρ on a set A is a subset of $A \times A$ with two elements $a, b \in A$ being *in relation* with respect to ρ if and only if $(a, b) \in \rho$, which we denote by $a \rho b$. If ρ is *reflexive* [$a \rho a \forall a \in A$], *antisymmetric* [$a \rho b$ and $b \rho a \Rightarrow a = b \forall a, b \in A$] and *transitive* [$a \rho b$ and $b \rho c \Rightarrow a \rho c \forall a, b, c \in A$], it is a *partial ordering relation* and A a *partially ordered set* or *poset*. We denote a poset A with partial ordering relation ρ by $\langle A; \rho \rangle$, abbreviated by simply A if the binary relation is clear.

For $H \subseteq A$ and $a \in A$, a is an *upper bound* of H if $h \leq a$ for all $h \in H$. a is the *least upper bound* or *supremum* of H if every upper bound b of H satisfies $a \leq b$ and we then write $a = \sup H$. *Lower bound* and *greatest lower bound* or *infimum*, denoted $\inf H$, are defined similarly. $\sup \emptyset$ is the smallest element in A , called *zero*, if it exists, and $\inf \emptyset$ is the largest element in A , called *unit*, if it exists.

A poset $\langle L; \leq \rangle$ is a *lattice* if $\inf H$ and $\sup H$ exist for any finite nonempty subset H of L . It is called *complete* if $\inf H$ and $\sup H$ also exist for $H = \emptyset$. A poset can be shown to be a complete lattice with the following result.

Lemma 2.1. *If $\langle P; \leq \rangle$ is a poset in which $\inf H$ exists for all $H \subseteq P$, then $\langle P; \leq \rangle$ is a complete lattice.*

For a lattice L and $a, b \in L$, we write $a \wedge b$ for $\inf\{a, b\}$ and $a \vee b$ for $\sup\{a, b\}$, and refer to \wedge as the *meet operation* and to \vee as the *join operation*. L is *distributive* if for all $a, b, c \in L$,

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \quad (4)$$

The structure of a lattice $\langle L; \leq \rangle$ may be visualised by a *Hasse diagram*, in which each element pair $a, b \in L$ is joined by an edge whenever $a \leq b$ and there is no $x \in L \setminus \{a, b\}$ such that $a \leq x \leq b$.

We denote most partial orderings by \leq . Which partial ordering the symbol refers to will be determined by the context. For an overview of lattice theory see Grätzer (1998).

3 Model Types: RCON and RCOR Models

3.1 RCON Models: Equality Restrictions on Concentrations

RCON models are graphical Gaussian models which place equality constraints on the entries of the concentration matrix $K = \Sigma^{-1}$. For a model whose conditional independence structure is represented by graph $G = (V, E)$, the restrictions can be represented by a graph colouring $(\mathcal{V}, \mathcal{E})$, with the vertex colouring \mathcal{V} representing constraints on the entries on the diagonal of K and the edge colouring \mathcal{E} representing constraints in the off-diagonal entries. Whenever two vertices $\alpha, \beta \in V$ belong to the same vertex colour class, the corresponding two diagonal entries $k_{\alpha\alpha}$ and $k_{\beta\beta}$ are restricted to being identical. Similarly, two edges $\alpha\beta, \gamma\delta \in E$ of the same colour represent the constraint $k_{\alpha\beta} = k_{\gamma\delta}$.

We denote the set of positive definite matrices which satisfy such constraints for a graph colouring $(\mathcal{V}, \mathcal{E})$ by $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$. Put formally, the distribution of a random vector $Y \in \mathbb{R}^V$ is said to lie in the RCON model represented by the coloured graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if

$$Y \sim \mathcal{N}_V(0, \Sigma), \quad K = \Sigma^{-1} \in \mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \left\{ \sum_{u \in \mathcal{V} \cup \mathcal{E}} \lambda_u T^u, \lambda \in \mathbb{R}^{\mathcal{V} \cup \mathcal{E}} \right\}^+ \quad (5)$$

Since the constraints are linear in K , by standard exponential family theory (Brown, 1986), just as unconstrained graphical Gaussian models, RCON models are regular exponential families. Thus the maximum likelihood estimate of λ is uniquely determined, provided it exists. For the corresponding computation algorithm see Højsgaard and Lauritzen (2008) and Højsgaard and Lauritzen (2007). Note that RCON models are instances of models considered in Anderson (1970).

Example 3.1. The data consist of the examination marks of 88 students in the mathematical subjects Algebra, Analysis, Mechanics, Statistics and Vectors (Mardia et al., 1979). Whittaker (1990); Edwards (2000) previously demonstrated an excellent fit to the unconstrained model represented by the graph shown in Figure 1(a). Højsgaard and Lauritzen (2008) show the data to also support the RCON model represented by the graph in Figure 1(b). The model specifications are

$$Y \sim \mathcal{N}_5(0, \Sigma), \quad \Sigma^{-1} \in M$$

with M given below the corresponding graphs. If the subjects are indexed in alphabetical order, the graph colouring representing the constraints of the RCON model is given by $(\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{\{1\}, \{2, 5\}, \{3, 4\}\}$ and $\mathcal{E} = \{\{12\}, \{13, 14, 15, 24, 35\}\}$. Note that the number of model parameters has been reduced from 11 to 5.

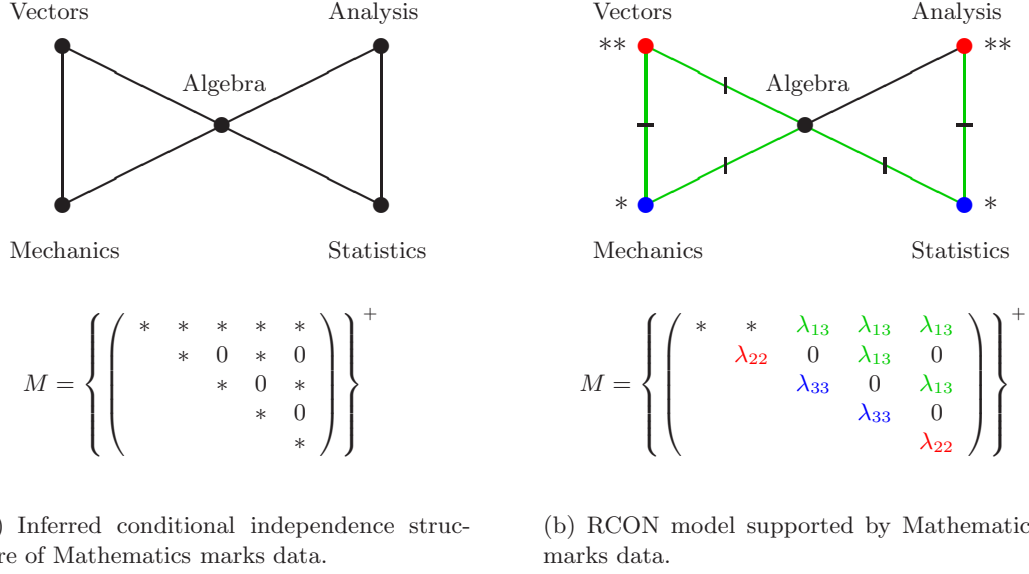


Figure 1: Mathematics marks example.

3.2 RCOR Models: Equality Restrictions on Partial Correlations

RCOR models place symmetry restrictions on the diagonal elements of the concentration matrix $K = \Sigma^{-1}$ and on the partial correlations as given in equation (2). Just as for RCON models, for a model with graph $G = (V, E)$, the constraints can be represented by a graph colouring $(\mathcal{V}, \mathcal{E})$: Vertices of the same colour represent restrictions on the diagonal entries of K (exactly as in RCON models), and whenever two edges $\alpha\beta, \gamma\delta \in E$ belong to the same edge colour class in \mathcal{E} , the corresponding partial correlations $\rho_{\alpha\beta|V \setminus \{\alpha, \beta\}}$ and $\rho_{\gamma\delta|V \setminus \{\gamma, \delta\}}$, defined in equation (2), are restricted to being identical.

We denote the set of positive definite matrices which satisfy such restrictions for a graph colouring $(\mathcal{V}, \mathcal{E})$ by $\mathcal{R}^+(\mathcal{V}, \mathcal{E})$. If we let the $|V| \times |V|$ matrices $A = (a_{\alpha\beta})$ and $C = (c_{\alpha\beta})$ be given by $a_{\alpha\beta} = \sqrt{k_{\alpha\beta}}$ for $\alpha = \beta$ and zero otherwise, and let $c_{\alpha\beta}$ as in equation (1) for $\alpha \neq \beta$ and $c_{\alpha\beta} = 1$ otherwise, then $K = ACA$ and the distribution of a random vector $Y \in \mathbb{R}^V$ lies in the RCOR model represented by the coloured graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if

$$Y \sim \mathcal{N}_V(0, \Sigma), \quad K = \Sigma^{-1} \in \mathcal{R}^+(\mathcal{V}, \mathcal{E}) = \left\{ ACA \mid A = \sum_{u \in \mathcal{V}} \eta_u T^u, \quad \eta \in \mathbb{R}_+^{\mathcal{V}}, \right. \\ \left. C = I + \sum_{u \in \mathcal{E}} \tau_u T^u, \quad \tau \in (-1, 1)^{\mathcal{E}} \right\}^+ \quad (6)$$

Thus the constraints of RCOR models define a differentiable manifold in \mathcal{S}^+ which makes them curved exponential families (Brown, 1986). Therefore, the maximum likelihood estimates of η and τ , if they exist, may not be unique. For a discussion and computation algorithm we refer to Højsgaard and Lauritzen (2008) and Højsgaard and Lauritzen (2007).

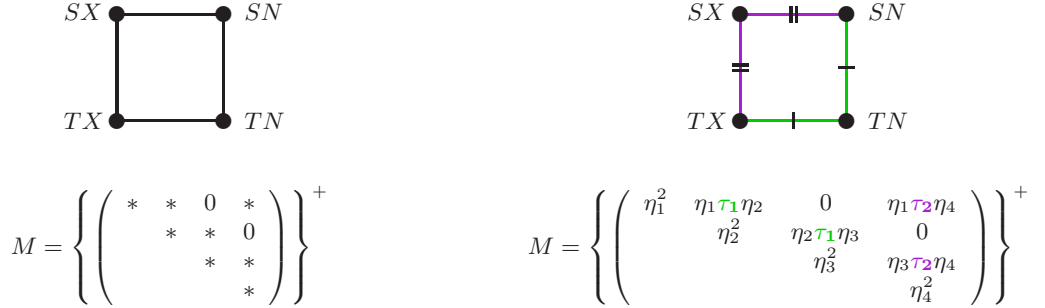
RCOR models are scale invariant if variables inside the same vertex colour class are manipulated in the same way (Højsgaard and Lauritzen, 2008). Thus they are particularly suitable for variables measured on different scales.

We highlight that both RCON and RCOR models generally do not place the same equality restrictions on Σ as they do on Σ^{-1} and on partial correlations.

Example 3.2. The data is concerned with anxiety and anger in a trait and state version of 684 students (Cox and Wermuth, 1993) and strongly support the conditional independence model displayed in Figure 2(a). As shown in Højsgaard and Lauritzen (2008), they also support the RCOR model represented by the coloured graph in Figure 2(b), parametrised by 6 parameters rather than 8. The variable names are combinations between T or S , for “trait” and “state”, and X or N , standing for “anxiety” and “anger”. The model specifications are

$$Y \sim \mathcal{N}_4(0, \Sigma), \quad \Sigma^{-1} \in M$$

with M given below the graphs. The variables are indexed anti-clockwise starting from TX .



(a) Conditional independence structure supported by personality characteristics data.

(b) RCOR model supported by personality characteristics data.

Figure 2: Personality characteristics example.

3.3 Number of RCON and RCOR Models

Let \mathcal{S}_V^+ and \mathcal{R}_V^+ denote the sets of RCON and RCOR models with variable set V and let \mathcal{C}_V be the set of vertex and edge coloured graphs with vertex set V . Further, let M_V be the set of unconstrained graphical Gaussian models with variable set V and U_V the set of undirected graphs with vertex set V .

As, by equations (5) and (6), for both model types there is one model parameter for each vertex colour class in \mathcal{V} and one for each edge colour class in \mathcal{E} in the coloured dependence graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, there are as many RCON and RCOR models with variables V as there are coloured graphs with vertex set V , *i.e.*, $|\mathcal{S}_V^+| = |\mathcal{R}_V^+| = |\mathcal{C}_V|$. Given that the number of graph colourings of a particular graph $G = (V, E)$ is given by the product $|P(V)||P(E)|$ of the number of partitions of V multiplied by the number of partitions of E , we obtain

$$|\mathcal{C}_V| = \sum_{G=(V,E) \in U_V} |P(V)||P(E)| = |P(V)| \sum_{G=(V,E) \in U_V} |P(E)| \quad (7)$$

For a discrete set S of size d , $|P(S)|$ is given the d^{th} Bell number B_d (Bell, 1934; Pitman, 1997), which satisfy the recursive relationship $B_{d+1} = \sum_{k=0}^d \binom{d}{k} B_k$, with $B_0 = 1$. Hence,

$$|\mathcal{S}_V^+| = |\mathcal{R}_V^+| = B_{|V|} \sum_{G=(V,E) \in U_V} B_{|E|} = B_{|V|} \sum_{k=0}^{\binom{|V|}{2}} \binom{\binom{|V|}{2}}{k} B_k = B_{|V|} B_{\binom{|V|}{2}+1}$$

For each d , B_d can be evaluated as the least integer greater than the sum of the first $2d$ terms in *Dobiński's formula* (Dobiński, 1877; Comlet, 1974)

$$B_d = e^{-1} \sum_{k=0}^{\infty} \frac{k^d}{k!} = e^{-1} \left(\frac{0^d}{0!} + \frac{1^d}{1!} + \frac{2^d}{2!} + \dots \right) \quad (8)$$

so that clearly $|\mathcal{S}_V^+| = |\mathcal{R}_V^+|$ grow super-exponentially in $|V|$. For illustration, observe that while $|M_{[4]}| = 64$ and $|M_{[5]}| = 1,024$, $|\mathcal{S}_{[4]}^+| = |\mathcal{R}_{[4]}^+| = 13,155$ and $|\mathcal{S}_{[5]}^+| = |\mathcal{R}_{[5]}^+| = 35,285,640$.

3.4 Structure of the Sets of RCON and RCOR Models

It is a well-known fact that M_V is a complete distributive lattice with respect to model inclusion, with partial ordering induced by the partial ordering on U_V given by edge set inclusion: for $G_1 = (V, E_1), G_2 = (V, E_2) \in U_V$, $G_1 \leq G_2$ whenever $E_1 \subseteq E_2$, with $G_1 \wedge G_2 = (V, E_1 \cap E_2)$ and $G_1 \vee G_2 = (V, E_1 \cup E_2)$. For $M_1, M_2 \in M_V$ represented by G_1, G_2 as above, $M_1 \subseteq M_2$ if and only if $G_1 \leq G_2$. The zero in $\langle U_V; \leq \rangle$ is the empty graph, and the unit the complete graph, in which every edge is present.

RCON and RCOR models are specified by partitions of V and E . For any finite discrete set S , the set $P(S)$ of partitions of S forms a complete non-distributive lattice, with $P_1 \leq P_2$ for $P_1, P_2 \in P(S)$ whenever P_1 is *finer* than P_2 , or, put differently, whenever P_2 is *coarser* than P_1 , *i.e.*, if every set in P_2 can be expressed as a union of sets in P_1 . This allows the identification of a partial ordering \preceq on \mathcal{C}_V which corresponds to model inclusion in \mathcal{S}_V^+ and \mathcal{R}_V^+ : For $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}}), \mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}}) \in \mathcal{C}_V$ with underlying uncoloured graphs G and H , $\mathcal{G} \preceq \mathcal{H}$ whenever

- (i) $G \leq H$; (ii) $\mathcal{V}_{\mathcal{G}} \geq \mathcal{V}_{\mathcal{H}}$; (iii) every colour class in $\mathcal{E}_{\mathcal{G}}$ is a union of colour classes in $\mathcal{E}_{\mathcal{H}}$.

Put in words, if we let $\mathcal{M}_{\mathcal{G}}, \mathcal{M}_{\mathcal{H}}$ denote two RCON or RCOR models (both of the same type) represented by $\mathcal{G}, \mathcal{H} \in \mathcal{C}_V$, then $\mathcal{M}_{\mathcal{G}} \subseteq \mathcal{M}_{\mathcal{H}}$ if \mathcal{H} can be obtained from \mathcal{G} by splitting of colour classes and adding new edge colour classes, or equivalently if \mathcal{G} can be obtained from \mathcal{H} by merging colour classes and dropping edge colour classes.

For example, for the graphs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ for $i = 1, 2, 3$ in Figure 3, $\mathcal{G}_1 \preceq \mathcal{G}_2$ as conditions (i)–(iii) above are satisfied whereas $\mathcal{G}_1 \not\preceq \mathcal{G}_3$ because (ii) and (iii) are violated. Thus the corresponding RCON or RCOR models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ (all of the same type) satisfy $\mathcal{M}_1 \subseteq \mathcal{M}_2$ and $\mathcal{M}_1 \not\subseteq \mathcal{M}_3$.

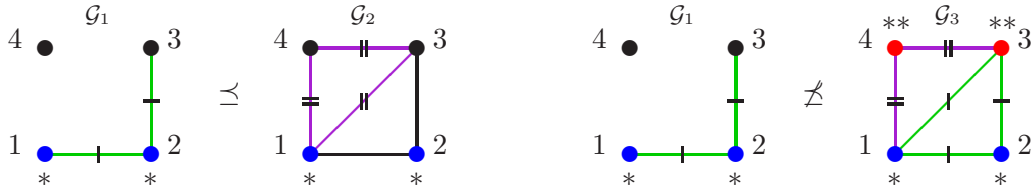


Figure 3: Partial ordering in $\mathcal{C}_{[4]}$.

It is then straight forward to show that $\langle \mathcal{C}_V; \preceq \rangle$ is a complete lattice with meet and join operations

$$\mathcal{G} \wedge \mathcal{H} = (\mathcal{V}_{\mathcal{G}} \vee \mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{G}}^* \vee \mathcal{E}_{\mathcal{H}}^*) \quad \text{and} \quad \mathcal{G} \vee \mathcal{H} = (\mathcal{V}_{\mathcal{G}} \wedge \mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{G}}^{**} \wedge \mathcal{E}_{\mathcal{H}}^{**}) \quad (9)$$

where $\mathcal{E}_G^* \subseteq \mathcal{E}_G$ and $\mathcal{E}_H^* \subseteq \mathcal{E}_H$ are maximal with the property that they are partitions of the same set of edges inside $E_G \cap E_H$, $\mathcal{E}_G^{**} = \mathcal{E}_G \cup \{\{E_H \setminus E_G\}\}$ and $\mathcal{E}_H^{**} = \mathcal{E}_H \cup \{\{E_G \setminus E_H\}\}$. The graphs in Figure 4 illustrate the operations. The zero in $\langle \mathcal{C}_V; \preceq \rangle$ is given by the empty graph in which all vertices are of the same colour and the unit is the complete graph with atomic colour classes.

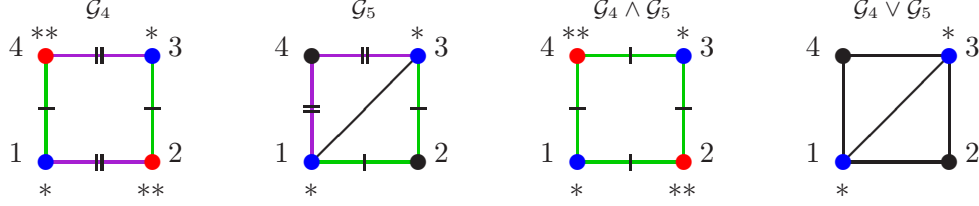


Figure 4: Meet and join operations in $\mathcal{C}_{[4]}$.

Proposition 3.3. *Let $\mathcal{G} = (\mathcal{V}_G, \mathcal{E}_G), \mathcal{H} = (\mathcal{V}_H, \mathcal{E}_H) \in \mathcal{C}_V$ and let $\mathcal{S}^+(\mathcal{V}_G, \mathcal{E}_G), \mathcal{S}^+(\mathcal{V}_H, \mathcal{E}_H) \in \mathcal{S}_V^+$ and $\mathcal{R}^+(\mathcal{V}_G, \mathcal{E}_G), \mathcal{R}^+(\mathcal{V}_H, \mathcal{E}_H) \in \mathcal{R}_V^+$ be the RCON and RCOR models represented by \mathcal{G} and \mathcal{H} . Then*

$$\mathcal{S}^+(\mathcal{V}_G, \mathcal{E}_G) \subseteq \mathcal{S}^+(\mathcal{V}_H, \mathcal{E}_H) \iff \mathcal{G} \preceq \mathcal{H} \iff \mathcal{R}^+(\mathcal{V}_G, \mathcal{E}_G) \subseteq \mathcal{R}^+(\mathcal{V}_H, \mathcal{E}_H)$$

and \mathcal{S}_V^+ and \mathcal{R}_V^+ are complete non-distributive lattices with meet and join operations induced by the meet and join operations in $\langle \mathcal{C}_V; \preceq \rangle$, given in equation (9).

4 Model Classes within RCON and RCOR Models

The motivation to study model classes strictly within the sets of RCON and RCOR models is three-fold: firstly, having demonstrated that the number of RCON and RCOR models grows dramatically with the number of variables, especially for model selection, smaller model (search) spaces are desirable. Secondly, generic equality constraints of RCON and RCOR models are generally not readily interpretable and, lastly, do not guarantee the corresponding model to have any particular statistical properties.

Four model classes within the sets of RCON and RCOR models which are characterised by desirable statistical properties expressing themselves in regularity of colouring have been previously identified in the literature. This section is devoted to their definition and first properties. Three of the four colouring regularities were termed *edge regularity*, with the corresponding models appearing in Højsgaard and Lauritzen (2008), *vertex regularity* and *regularity* in Gehrmann and Lauritzen (2011). We term colourings of the fourth type *permutation-generated*. The corresponding models are referred to as *graphical symmetry models* in Hylleberg et al. (1993) and as *RCOP models* in Højsgaard and Lauritzen (2008).

4.1 Models Represented by Edge Regular Colourings

RCON and RCOR models place restrictions on different parameter sets, which translates into different model properties. While the restrictions in RCON models ensure the models to be regular exponential families, RCOR models are scale invariant within vertex colour classes. Thus if a graph colouring $(\mathcal{V}, \mathcal{E})$ yields the same model restrictions representing the constraints

of an RCON model as it does representing those of an RCOR model, it represents a model with both of the above desirable properties. Such models can be identified by their graph colouring.

Definition 4.1. For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ we say that $(\mathcal{V}, \mathcal{E})$ is *edge regular* if any pair of edges in the same edge colour class in \mathcal{E} connects the same vertex colour classes in \mathcal{V} .

It then holds:

Proposition 4.2 (Højsgaard and Lauritzen (2008)). *The RCON and RCOR models determined by $(\mathcal{V}, \mathcal{E})$ yield identical restrictions*

$$\mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \mathcal{R}^+(\mathcal{V}, \mathcal{E})$$

if and only if $(\mathcal{V}, \mathcal{E})$ is edge regular.

We provide a simple example for illustration. While the colouring in Figure 5(a) is edge regular (both green edges (single dash) connect a blue vertex (single asterisk) to a red one (two asterisks), and the same is true for the purple edges (two dashes)), the colouring in Figure 5(b) is not, as the green edges appear between different pairs of vertex colours.

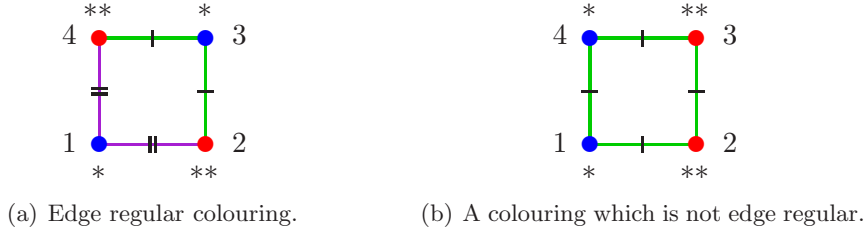


Figure 5: An edge regular colouring and one which is not edge regular.

4.2 Models Represented by Vertex Regular Colourings

Vertex regular colourings are of relevance to the estimation of a non-zero mean vector μ in a $\mathcal{N}_{|V|}(\mu, \Sigma)$ distribution if μ is subject to equality constraints and Σ^{-1} is restricted to lie inside $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ or inside $\mathcal{R}^+(\mathcal{V}, \mathcal{E})$ for some coloured graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Proposition 4.3 (Gehrmann and Lauritzen (2011)). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ and let \mathcal{M} be a partition of V . For $\alpha \in V$ let v_α denote the set in \mathcal{M} which contains α and let*

$$\Omega = \Omega(\mathcal{M}) = \{(x_\alpha)_{\alpha \in V} \in \mathbb{R}^V : x_\alpha = x_\beta \text{ whenever } \alpha \equiv \beta \ (\mathcal{M})\}$$

Further let $(Y^i)_{1 \leq i \leq n}$ be a sample of independent identically distributed observations $Y^i \sim \mathcal{N}_{|V|}(\mu, \Sigma)$ with μ restricted to lie inside Ω . Then the following are equivalent

- (i) *the likelihood function based on $(y^i)_{1 \leq i \leq n}$ is maximised in μ by the least-squares estimator μ^* for all Σ with $\Sigma^{-1} \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$ or with $\Sigma^{-1} \in \mathcal{R}^+(\mathcal{V}, \mathcal{E})$ where*

$$\mu_\alpha^* = \frac{\sum_{i=1}^n \sum_{\beta \in v_\alpha} y_\beta^i}{|v_\alpha|n} \tag{10}$$

- (ii) *\mathcal{M} is finer than \mathcal{V} and $(\mathcal{M}, \mathcal{E})$ is vertex regular.*

For the definition of a vertex regular colouring we require the concept of an *equitable partition*, first defined in Sachs (1966). For an undirected graph $G = (V, E)$, a vertex colouring \mathcal{V} of V is called *equitable with respect to G* if for all $v, w \in \mathcal{V}$ and all $\alpha, \beta \in v$, we have $|\text{ne}(\alpha) \cap w| = |\text{ne}(\beta) \cap w|$. *Vertex regular* graph colourings are the analogue to equitable partitions for vertex and edge coloured graphs.

Definition 4.4. For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ let the subgraph induced by the edge colour class $u \in \mathcal{E}$ be denoted by $G^u = (V, u)$. We say that $(\mathcal{V}, \mathcal{E})$ is *vertex regular* if \mathcal{V} is equitable with respect to G^u for all $u \in \mathcal{E}$.

While the colouring in Figure 6(a) is vertex regular, the colouring in Figure 6(b) is not. The former has only one edge colour class, so that it is vertex regular if and only if its vertex colouring is equitable with respect to G , which it is. The colouring on the right cannot be vertex regular as while vertex 4 is incident to a purple edge (two dashes), vertex 2 isn't even though they are of the same colour.

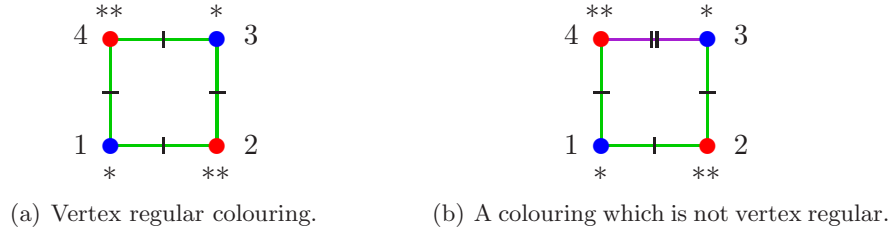


Figure 6: A vertex regular colouring and one which is not vertex regular.

4.3 Models Represented by Regular Colourings

RCON or RCOR models with restrictions represented by colourings which are both edge regular and vertex regular combine the properties of both model classes. It can be shown that the colourings of such models are precisely those which in the terminology of Siemons (1983) are *regular*:

Definition 4.5 (Siemons (1983)). For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$, $(\mathcal{V}, \mathcal{E})$ is *regular* if

- (i) every pair of equally coloured edges in \mathcal{E} connects the same vertex colour classes in \mathcal{V} ;
- (ii) every pair of equally coloured vertices in \mathcal{V} has the same degree in every edge colour class in \mathcal{E} .

By the above, the colourings shown in Figure 5(b) and Figure 6(b) cannot be regular. While the colouring given in Figure 6(a) is regular, the colouring in Figure 5(a) is not.

4.4 Models Represented by Permutation-Generated Colourings

Permutation-generated colourings are a special instance of regular colourings (for a proof see Gehrmann and Lauritzen (2011)), and thus by definition also of edge regular and vertex regular colourings. They represent models in which equality constraints on the parameters are induced by permutation symmetry and allow a particularly simple maximisation of the likelihood function. In brief, maximum likelihood estimates can be obtained by standard methods

for unconstrained models after taking averages within colour classes Højsgaard and Lauritzen (2008). Further, models represented by permutation-generated colourings form the only model class discussed here which restricts Σ^{-1} and Σ in the same fashion.

The corresponding models are defined through distribution invariance under a permutation group Γ acting on the variable labels V . If $Y \sim \mathcal{N}_{|V|}(0, \Sigma)$, then permutations acting on V simultaneously permute rows and columns of Σ so that the distribution of Y is invariant under $\Gamma \subseteq S(V)$ if and only if

$$P(\sigma)\Sigma P(\sigma)^T = \Sigma \iff P(\sigma)\Sigma = \Sigma P(\sigma) \iff P(\sigma)\Sigma^{-1} = \Sigma^{-1}P(\sigma) \quad (11)$$

for all $\sigma \in \Gamma$, where for $\alpha, \beta \in V$, $P(\sigma)_{\alpha\beta} = 1$ if and only if σ maps β to α and zero otherwise. A necessary condition for equation (11) to hold is that the zero entries in Σ^{-1} are preserved for all Σ in the model and all $\sigma \in \Gamma$. Thus if the distribution of Y is assumed to lie in the graphical Gaussian model represented by graph $G = (V, E)$, by equation (3), we in particular require that $\Gamma \subseteq \text{Aut}(G)$.

Therefore, in the notation in Højsgaard and Lauritzen (2008), a graphical Gaussian model with conditional independence structure represented by graph G which is permutation invariant under group $\Gamma \subseteq \text{Aut}(G)$ is given by assuming

$$\Sigma^{-1} \in \mathcal{S}^+(G) \cap \mathcal{S}^+(\Gamma)$$

where $\mathcal{S}^+(\Gamma)$ is the set of positive definite matrices Σ satisfying the equivalent conditions in equation (11).

By definition, permutation invariant models place constraints on all model parameters and thus in particular on concentrations and partial correlations, which they restrict in the same fashion. Thus symmetry constraints in permutation invariant models can be represented by a vertex and edge colouring $(\mathcal{V}, \mathcal{E})$ of G given by the orbits of Γ in V and E respectively, *i.e.*, by giving two vertices $\alpha, \beta \in V$ the same colour whenever there exists $\sigma \in \Gamma$ mapping α to β , and similarly for the edges. We term such colourings *permutation-generated*, formally defined below.

Definition 4.6. For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ with underlying uncoloured graph $G = (V, E)$ we say that $(\mathcal{V}, \mathcal{E})$ is *permutation-generated* if there exists a group $\Gamma \subseteq \text{Aut}(G)$ acting on V such that \mathcal{V} and \mathcal{E} are given by the orbits of Γ in V and E respectively.

The following example illustrates that in addition to the aforementioned desirable statistical properties, models with permutation-generated restrictions allow a very intuitive interpretation.

Example 4.7. The data, commonly referred to as *Fret's heads*, is concerned with the head dimensions of 25 pairs of first and second sons (Frets, 1921; Mardia et al., 1979). Previous analyses (Whittaker, 1990) support a model represented by the graph in Figure 7(a), where L_i and B_i denote the head length and head breadth of son i for $i = 1, 2$. Højsgaard and Lauritzen (2008) showed the model generated by $\Gamma = \langle (B_1 B_2)(L_1 L_2) \rangle$, corresponding to permuting the two sons, represented by the first graph in Figure 7(b) to be an excellent fit.

Another model with constraints generated by permutation symmetry which fits the data very well is the complete symmetry model, generated by $\Gamma = S(V)$, which is represented by the second coloured graph in Figure 7(b). Interestingly, it is further favourable over the former with regards to parameter estimation. While the graph on the left in Figure 7(b) is

non-decomposable and symmetry arguments combined with results in Buhl (1993) give that at least 2 observations are required for almost sure existence of $\hat{\Sigma}$, see also Uhler (2010), the complete symmetry model only requires one observation for $\hat{\Sigma}$ to exist almost surely.

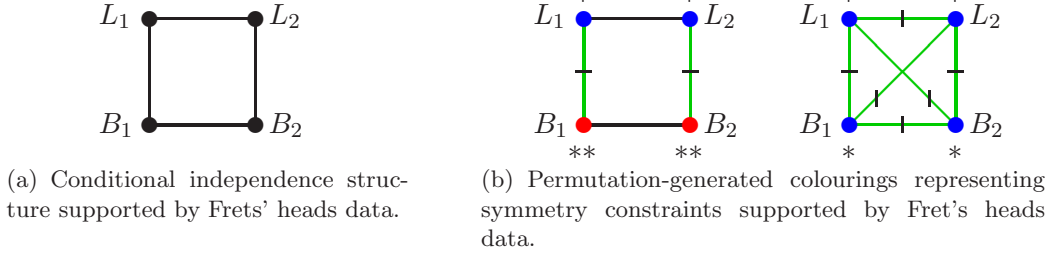


Figure 7: Frets' heads example.

4.5 Relations Between Model Classes

Let B , P , R and Π denote the sets of edge regular, vertex regular, regular and permutation-generated colourings respectively. The structural relations between colouring classes are summarised in the diagram displayed in Figure 8. In fact, we already saw examples of colourings in three of the four disjoint sets in the diagram. The colouring displayed in Figure 5(b) lies in $P \setminus R$, Figure 7(b) shows a colouring in Π and the colouring in Figure 6(b) lies in $B \setminus R$. (A graph colouring in $R \setminus \Pi$ is given by $(\mathcal{V}, \mathcal{E})$ with $V = [11]$, $\mathcal{V} = \{\{1, 2, 3\}, \{4, 5, 6, 7, 8, 9\}, \{10, 11\}\}$ and $\mathcal{E} = \{\{14, 15, 26, 27, 38, 39\}, \{(4, 10), (5, 10), (6, 10), (7, 11), (8, 11), (9, 11)\}\}$ where (i, j) denotes an edge between vertices i and j .)

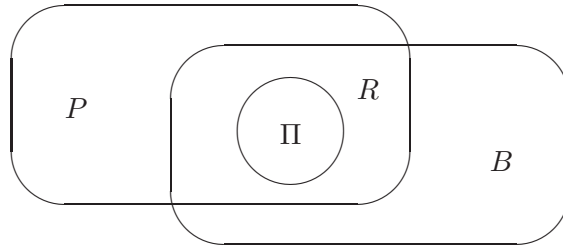


Figure 8: Structural relations between colouring classes.

By Proposition 4.2, a graph colouring yields the same model restrictions representing an RCON model as it does representing an RCOR model if and only if it lies in B . Therefore, the model type only needs to be specified whenever a graph colouring lies in $P \setminus B$. Put formally:

Proposition 4.8. *For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$, $\mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \mathcal{R}^+(\mathcal{V}, \mathcal{E})$ for $(\mathcal{V}, \mathcal{E}) \in B$ and $\mathcal{S}^+(\mathcal{V}, \mathcal{E}) \neq \mathcal{R}^+(\mathcal{V}, \mathcal{E})$ for $(\mathcal{V}, \mathcal{E}) \in P \setminus B$. Thus if for $X \in \{B, P, R, \Pi\}$ we let \mathcal{S}_X^+ denote the set of RCON models represented by graphs in X and similarly for \mathcal{R}_X^+ and RCOR models, then*

$$\mathcal{S}_B^+ = \mathcal{R}_B^+, \quad \mathcal{S}_R^+ = \mathcal{R}_R^+, \quad \mathcal{S}_\Pi^+ = \mathcal{R}_\Pi^+, \quad \mathcal{S}_P^+ \neq \mathcal{R}_P^+$$

giving rise to five model classes lying strictly within \mathcal{S}_V^+ and \mathcal{R}_V^+ with

$$\mathcal{S}_\Pi^+ \subset \mathcal{S}_R^+ \subset \mathcal{S}_B^+, \quad \mathcal{S}_B^+ \cap \mathcal{S}_P^+ = \mathcal{S}_B^+ \cap \mathcal{R}_P^+ = \mathcal{S}_R^+.$$

Let B_V , P_V , R_V and Π_V denote the sets of graph colourings inside B , P , R and Π with vertex set V . By Proposition 4.8, there are five corresponding model classes: $\mathcal{S}_{B_V}^+$, $\mathcal{S}_{P_V}^+$, $\mathcal{R}_{P_V}^+$, $\mathcal{S}_{R_V}^+$ and $\mathcal{S}_{\Pi_V}^+$. For illustration we give the corresponding model class sizes for $V = [4]$ together with $|M_{[4]}|$ in Table 1.

Model class	$\mathcal{S}_{[4]}^+, \mathcal{R}_{[4]}^+$	$\mathcal{S}_{B_{[4]}}^+$	$\mathcal{S}_{P_{[4]}}^+, \mathcal{R}_{P_{[4]}}^+$	$\mathcal{S}_{R_{[4]}}^+$	$\mathcal{S}_{\Pi_{[4]}}^+$	$M_{[4]}$
Size	13,155	3065	1380	251	251	64

Table 1: Model set sizes for $V = [4]$.

The relative sizes in Table 1 are representative of the general case: $\mathcal{S}_{B_V}^+$ is the largest model class, followed by $\mathcal{S}_{P_V}^+$, $\mathcal{R}_{P_V}^+$ and $\mathcal{S}_{R_V}^+$. $|\mathcal{S}_{P_V}^+| = |\mathcal{R}_{P_V}^+|$ will generally be considerably smaller than $\mathcal{S}_{B_V}^+$ as the defining conditions of P_V are far more restrictive than those for B_V . $|\mathcal{S}_{\Pi_V}^+|$ is the smallest class of the four for all V , however may equal $|\mathcal{S}_{R_V}^+|$ for some V , as for example for $V = [4]$.

5 Structures of Model Classes

Below we show that each model class defined above forms a complete non-distributive lattice, starting with $\mathcal{S}_{B_V}^+$ and $\mathcal{S}_{\Pi_V}^+$ as their structure turns out most tractable. For brevity, we only outline results for the remaining model classes.

5.1 Models Represented by Edge Regular Colourings

Proposition 5.1. *B_V is stable under the meet operation \wedge in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9).*

Proof: Let $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}}), \mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}}) \in B_V$. $\mathcal{G} \wedge \mathcal{H} = (\mathcal{V}_{\mathcal{G} \wedge \mathcal{H}}, \mathcal{E}_{\mathcal{G} \wedge \mathcal{H}})$ is obtained from \mathcal{G} and \mathcal{H} by dropping of edge colour classes and merging of colour classes. The only operation potentially leading to $\mathcal{G} \wedge \mathcal{H}$ lying outside of B_V is merging of edge colour classes.

So let $\alpha\beta$ and $\gamma\delta$ be two edges in $\mathcal{G} \wedge \mathcal{H}$ of equal colour. Then there exists a sequence $\alpha_0\beta_0, \dots, \alpha_k\beta_k$ in $E_{\mathcal{G} \wedge \mathcal{H}}$ such that $\alpha_0\beta_0 = \alpha\beta$, $\alpha_k\beta_k = \gamma\delta$, and $\alpha_{i-1}\beta_{i-1} \equiv \alpha_i\beta_i$ ($\mathcal{E}_{\mathcal{G}}$) or $\alpha_{i-1}\beta_{i-1} \equiv \alpha_i\beta_i$ ($\mathcal{E}_{\mathcal{H}}$) for $1 \leq i \leq k$. As both \mathcal{G} and \mathcal{H} have edge regular colourings, $\alpha_{i-1}\beta_{i-1}$ and $\alpha_i\beta_i$ connect the same vertex colour classes in the graph in which they are of equal colour, which we denote by $\{\alpha_{i-1}, \beta_{i-1}\} \equiv \{\alpha_i, \beta_i\}$ ($\mathcal{V}_{\mathcal{X}}$) with $\mathcal{X} \in \{\mathcal{G}, \mathcal{H}\}$. This gives that $\{\alpha, \beta\} = \{\alpha_0, \beta_0\} \equiv \{\alpha_k, \beta_k\} = \{\gamma, \delta\}$ ($\mathcal{V}_{\mathcal{G} \vee \mathcal{H}}$). As $\mathcal{V}_{\mathcal{G}} \vee \mathcal{V}_{\mathcal{H}} = \mathcal{V}_{\mathcal{G} \wedge \mathcal{H}}$, $\alpha\beta$ and $\gamma\delta$ connect the same vertex colour classes in $\mathcal{G} \wedge \mathcal{H}$. \square

Graphs \mathcal{G}_4 , \mathcal{G}_5 and $\mathcal{G}_4 \wedge \mathcal{G}_5$ in Figure 4 illustrate the stability of B_V under \wedge . That B_V is generally not stable under the join operation \vee in equation (9) is established by the example in Figure 9.

Proposition 5.2. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ and let \mathcal{E}_B be the partition of E which puts $\alpha\beta, \gamma\delta \in E$ in the same set whenever they connect the same vertex color classes. Then \mathcal{G} has a supremum in B_V , given by*

$$s_B(\mathcal{G}) = (\mathcal{V}, \mathcal{E} \wedge \mathcal{E}_B)$$

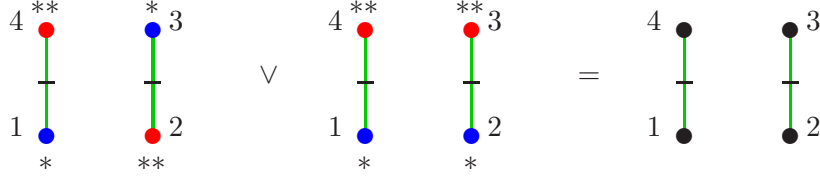


Figure 9: $B_{[4]}$ is not stable under \vee .

Proof: The claim is trivially true for $\mathcal{G} \in B_V$. If $\mathcal{G} \in \mathcal{C}_V \setminus B_V$, an edge regular colouring cannot be achieved through splitting vertex colour classes or adding edge colour classes. The only effective manipulation is therefore the splitting of edge colour classes. The coarsest partition which is finer than \mathcal{E} and gives an edge regular colouring is $\mathcal{E} \wedge \mathcal{E}_B$, as it splits edge colour classes only if they connect different vertex colour classes. \square

We deduce:

Theorem 5.3. $\mathcal{S}_{B_V}^+$ is a complete non-distributive lattice with respect to model inclusion. The meet operation is induced by the meet operation in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9). The join of two models represented by graphs $\mathcal{G}, \mathcal{H} \in B_V$ is represented by the graph $s_B(\mathcal{G} \vee \mathcal{H})$.

Proof: Proposition 5.1 implies that $\inf H$ exists for all finite $H \subseteq B_V$, which by Lemma 2.1 gives that B_V is a complete lattice, with the same meet operation as $\langle \mathcal{C}_V; \preceq \rangle$. As the zero and unit in $\langle \mathcal{C}_V; \preceq \rangle$ have edge regular colourings, they are also the zero and unit in B_V .

By definition, for $\mathcal{G}, \mathcal{H} \in B_V$, the join \mathcal{K} of \mathcal{G} and \mathcal{H} in B_V is the smallest graph with respect to partial ordering \preceq which satisfies $\mathcal{K} \succeq \mathcal{G}$, $\mathcal{K} \succeq \mathcal{H}$ and $\mathcal{K} \in B_V$. The supremum $\mathcal{G} \vee \mathcal{H}$ of \mathcal{G} and \mathcal{H} in $\langle \mathcal{C}_V; \preceq \rangle$ is the smallest graph satisfying the first two relations so that \mathcal{K} is in fact the smallest graph satisfying $\mathcal{K} \succeq (\mathcal{G} \vee \mathcal{H})$ and $\mathcal{K} \in B_V$. Thus $\mathcal{G} \vee_B \mathcal{H}$ is given by the supremum of $\mathcal{G} \vee \mathcal{H}$ in B_V , which by Proposition 5.2 equals $s_B(\mathcal{G} \vee \mathcal{H})$.

Non-distributivity of B_V is established by observing that equation (4) is violated for $a = \mathcal{G}_6$, $b = \mathcal{G}_7$ and $c = \mathcal{G}_8$ displayed in Figure 10, as $\mathcal{G}_6 = \mathcal{G}_6 \vee (\mathcal{G}_7 \wedge \mathcal{G}_8) \neq (\mathcal{G}_6 \vee \mathcal{G}_7) \wedge (\mathcal{G}_6 \vee \mathcal{G}_8) = \mathcal{G}_6 \vee \mathcal{G}_7$.

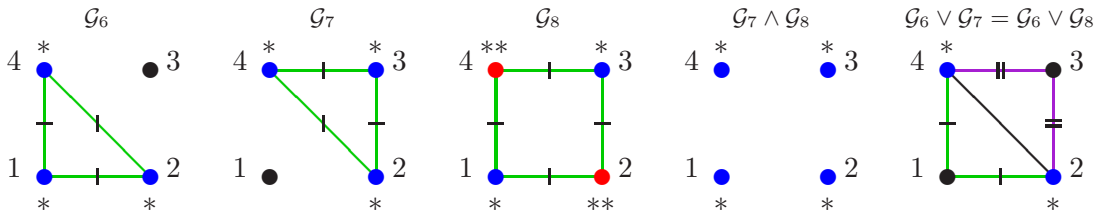


Figure 10: Non-distributivity of B_V .

The results on the structure of B_V naturally translate to the set of models $\mathcal{S}_{B_V}^+$, proving the claim. \square

5.2 Models Represented by Permutation-Generated Colourings

Let Γ_V denote the set of permutation groups acting on V . Then $\langle \Gamma_V; \subseteq \rangle$ is a complete lattice Schmidt (1994) with meet and join operations given by $\Gamma_1 \wedge \Gamma_2 = \Gamma_1 \cap \Gamma_2$ and $\Gamma_1 \vee \Gamma_2 = \langle \Gamma_1 \cup \Gamma_2 \rangle$. We obtain:

Proposition 5.4. Π_V is stable under the meet operation \wedge in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9). If $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}}), \mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}}) \in \Pi_V$ are generated by $\Gamma_{\mathcal{G}}, \Gamma_{\mathcal{H}} \in \Gamma_V$, then the colouring of $\mathcal{G} \wedge \mathcal{H}$ is generated by $\Gamma_{\mathcal{G}} \vee \Gamma_{\mathcal{H}}$.

Proof: Let $\mathcal{G}, \mathcal{H} \in \Pi_V$ and $\Gamma_{\mathcal{G}}, \Gamma_{\mathcal{H}} \in \Gamma_V$ be as in the claim. Then $\mathcal{V}_{\mathcal{G}}$ and $\mathcal{E}_{\mathcal{G}}$ are unions of orbits of $\Gamma_{\mathcal{G}}$ in V and $E_{\mathcal{G}}$, and similarly for $\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}}$ and $\Gamma_{\mathcal{H}}$. By definition of the meet operation in $\langle \mathcal{C}_V; \preceq \rangle$, each vertex colour class in $\mathcal{G} \wedge \mathcal{H} = (\mathcal{V}_{\mathcal{G} \wedge \mathcal{H}}, \mathcal{E}_{\mathcal{G} \wedge \mathcal{H}})$ can be expressed as a union of vertex colour classes in $\mathcal{V}_{\mathcal{G}}$, and as a union of vertex colour classes in $\mathcal{V}_{\mathcal{H}}$, and similarly for the edges. Thus the colouring of $\mathcal{G} \wedge \mathcal{H}$ is invariant under the action of both groups $\Gamma_{\mathcal{G}}$ and $\Gamma_{\mathcal{H}}$, and therefore also under $\Gamma_{\mathcal{G}} \vee \Gamma_{\mathcal{H}}$.

To show that the colouring of $\mathcal{G} \wedge \mathcal{H}$ is generated by $\Gamma_{\mathcal{G}} \vee \Gamma_{\mathcal{H}}$, we need to show that whenever two vertices or edges in $\mathcal{G} \wedge \mathcal{H}$ are of the same colour, then there exists $\sigma \in \Gamma_{\mathcal{G}} \vee \Gamma_{\mathcal{H}}$ which maps one of them to the other. We present the argument for the edges only, as it can be trivially transferred to the vertices. So let $\alpha\beta$ and $\gamma\delta$ be two edges in $\mathcal{G} \wedge \mathcal{H}$ of equal colour. Then there exists a sequence $\alpha_0\beta_0, \dots, \alpha_k\beta_k$ in $E_{\mathcal{G} \wedge \mathcal{H}}$ such that $\alpha_0\beta_0 = \alpha\beta$, $\alpha_k\beta_k = \gamma\delta$, and $\alpha_{i-1}\beta_{i-1} \equiv \alpha_i\beta_i$ ($\mathcal{E}_{\mathcal{G}}$) or $\alpha_{i-1}\beta_{i-1} \equiv \alpha_i\beta_i$ ($\mathcal{E}_{\mathcal{H}}$) for $1 \leq i \leq k$. There must therefore exist $\sigma_i \in \Gamma_{\mathcal{G}} \cup \Gamma_{\mathcal{H}}$ such that $\alpha_{i-1}\beta_{i-1}$ is mapped to $\alpha_i\beta_i$ by σ_i for $1 \leq i \leq k$, giving that the product $\sigma_k \dots \sigma_1 \in \Gamma_{\mathcal{G}} \vee \Gamma_{\mathcal{H}}$ maps $\alpha\beta$ to $\gamma\delta$. \square

Graphs $\mathcal{G}_4, \mathcal{G}_5$ and $\mathcal{G}_4 \wedge \mathcal{G}_5$ in Figure 4 illustrate the above result, as each of the graphs lies in Π_V , with generating groups $\Gamma_4 = \langle (13)(24) \rangle$, $\Gamma_5 = \langle (13) \rangle$ and $\Gamma_{4 \wedge 5} = \Gamma_4 \vee \Gamma_5 = \langle (13), (24) \rangle$ respectively. Observing that the join $\mathcal{G}_4 \vee \mathcal{G}_5$, also displayed in Figure 4, does not lie in Π_V establishes that Π_V is generally not stable under the join operation in $\langle \mathcal{C}_V; \preceq \rangle$. However the following holds:

Proposition 5.5. For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$, let $\text{Aut}(\mathcal{V}, \mathcal{E}) \leq S(V)$ denote the largest group leaving $(\mathcal{V}, \mathcal{E})$ invariant and let $(\mathcal{V}_{\text{Aut}}, \mathcal{E}_{\text{Aut}})$ denote the graph colouring of $G = (V, E)$ given by the orbits of $\text{Aut}(\mathcal{V}, \mathcal{E})$ in V and E respectively. Then \mathcal{G} has a supremum in Π_V given by

$$s_{\Pi}(\mathcal{G}) = (\mathcal{V}_{\text{Aut}}, \mathcal{E}_{\text{Aut}})$$

Proof: The claim is trivially true if $\mathcal{G} \in \Pi_V$. So suppose $\mathcal{G} \in \mathcal{C}_V \setminus \Pi_V$. \mathcal{G} is modified to a larger graph by adding edge colour classes and splitting colour classes. As the former will not enforce a permutation-generated colouring, to prove the claim we need to that $(\mathcal{V}_{\text{Aut}}, \mathcal{E}_{\text{Aut}})$ is the coarsest refinement of $(\mathcal{V}, \mathcal{E})$ which lies in Π_V . This is clearly the case as $\text{Aut}(\mathcal{V}, \mathcal{E})$ is the largest group which leaves \mathcal{V} and \mathcal{E} invariant. \square

Theorem 5.6. $\mathcal{S}_{\Pi_V}^+$ is a complete non-distributive lattice with respect to model inclusion. The meet operation is induced by the meet operation in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9). The join of two models represented by graphs $\mathcal{G}, \mathcal{H} \in \Pi_V$ is represented by the graph $s_{\Pi}(\mathcal{G} \vee \mathcal{H})$.

Proof: The proof is analogous to the proof of Theorem 5.3. In brief, Proposition 5.4 and Lemma 2.1 give that Π_V is a complete lattice with meet operation as claimed. As the zero and unit in $\langle \mathcal{C}_V; \preceq \rangle$ have permutation-generated colourings, with $\Gamma_0 = S(V)$ and $\Gamma_1 = \langle Id \rangle$, they are the zero and unit in Π_V .

By Proposition 5.5, the join of two graphs $\mathcal{G}, \mathcal{H} \in \Pi_V$ is given by $s_{\Pi}(\mathcal{G} \vee \mathcal{H})$. The graphs displayed in Figure 10 establish non-distributivity of Π_V as each of them has a permutation-generated colouring, with $\Gamma_6 = \langle (124) \rangle$, $\Gamma_7 = \langle (234) \rangle$, $\Gamma_8 = \langle (13), (24) \rangle$, $\Gamma_{7 \wedge 8} = S([4])$ and $\Gamma_{6 \vee 7} = \Gamma_{6 \vee 8} = \langle (24) \rangle$ respectively. The above directly translate to $\mathcal{S}_{\Pi_V}^+$, proving the claim. \square

5.3 Models Represented by Regular and Vertex Regular Colourings

The structures of R_V and P_V turn out to be closely related and it is for that reason that we treat them together. We abstain from giving explicit proofs of intermediate results for brevity, however give all facts the reader will require to construct them. We shall employ the notion of a factor graph:

Definition 5.7 (Frey (1998)). Let $f(Y)$ be a function in Y which factorises as $f(Y) = \prod_{i \in I} f_{A_i}(Y_{A_i})$ where $A_i \subseteq V$ and f_{A_i} cannot be factorised further for $i \in I$. Then the *factor graph* of f is the graph $G_F = (V \cup F, E_F)$ with $F = \{f_{A_i}\}_{i \in I}$ being the set of factor vertices and $E_F = \{\alpha f_{A_i} \mid \alpha \in V, f_{A_i} \in F \text{ with } \alpha \in A_i\}$.

For $Y = (Y_\alpha)_{\alpha \in V}$ assumed to follow a $\mathcal{N}_{|V|}(\mu, \Sigma)$ distribution, the density $f(y)$ factorises as

$$f(y) \propto \prod_{\alpha \in V} \exp\{-k_{\alpha\alpha}(y_\alpha - \mu_\alpha)^2/2\} \cdot \prod_{\substack{\alpha, \beta \in V, \\ \alpha \neq \beta}} \exp\{-k_{\alpha\beta}(y_\alpha - \mu_\alpha)(y_\beta - \mu_\beta)\}$$

giving that for the Gaussian distribution, either $A_i = \{\alpha\}$ or $A_i = \{\alpha, \beta\}$ for $\alpha, \beta \in V$, with a factor being present if and only if the corresponding entry in Σ^{-1} is non-zero. Thus if the distribution of Y is assumed lie in the graphical Gaussian model represented by graph $G = (V, E)$, by equation (3), each factor corresponds to a vertex in V or edge in E . The vertices in V can clearly be identified with their factors so that the factor graph of a graphical Gaussian model with graph $G = (V, E)$ equals

$$G_F = (V \cup F, E_F) \text{ with } F = \{e \mid e \in E\} \text{ and } E_F = \{\alpha e \mid \alpha \in V, e \in E \text{ is incident with } \alpha \text{ in } G\}.$$

This can be extended to the notion of a coloured factor graph, with an example given in Figure 11.

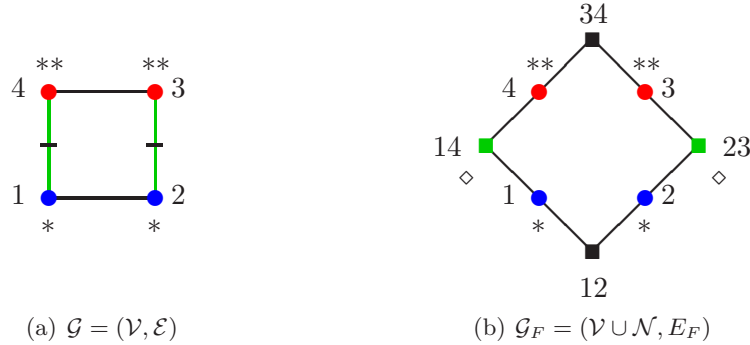


Figure 11: A coloured graph and the corresponding coloured factor graph.

Definition 5.8. For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ representing a graphical Gaussian model with equality constraints, let N be a set of nodes with each node representing an edge in E and let \mathcal{N} be the colouring of N in which nodes receive the same colour if and only if the corresponding edges are equally coloured in \mathcal{E} . The *coloured factor graph* of the model is defined to be the vertex and node coloured graph $\mathcal{G}_F = (\mathcal{V} \cup \mathcal{N}, E_F)$ with $E_F = \{\alpha n \mid \alpha \in V, n \in N \text{ and } n \text{ represents an edge incident with } \alpha\}$. The set of coloured factor graphs with vertex set V is denoted \mathcal{F}_V .

We give four intermediate results whose proofs we omit for brevity.

Lemma 5.9. \mathcal{F}_V and \mathcal{C}_V are isomorphic. We denote the isomorphism by $\phi_V : \mathcal{C}_V \rightarrow \mathcal{F}_V$.

Lemma 5.10. If $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ and $\phi_V(\mathcal{G}) = \mathcal{G}_F = (\mathcal{V} \cup \mathcal{N}, E_F)$ is the corresponding factor graph, then $(\mathcal{V}, \mathcal{E}) \in R_V$ if and only if $(\mathcal{V} \cup \mathcal{N})$ is equitable with respect to $G_F = (V \cup N, E_F)$.

Lemma 5.11 (McKay (1976)). If P_1 and P_2 are two partitions of V both equitable with respect to the same graph $G = (V, E)$, then so is their join $P_1 \vee P_2$.

Lemma 5.12 (McKay (1976)). For $G = (V, E)$ and a partition P of V , (up to the order of cells) there exists a unique coarsest partition that is finer than P and equitable with respect to G , to be denoted by $r_G(P)$.

Combined, Lemmas 5.9, 5.10 and 5.11 can be used to prove:

Proposition 5.13. R_V is stable under the meet operation \wedge in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9).

Note that the graphs in Figure 4 illustrate the stability of R_V under \wedge while showing that R_V is generally not stable under \vee in $\langle \mathcal{C}_V; \preceq \rangle$. Further, Lemma 5.12 implies:

Proposition 5.14. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ and let $\phi_V(\mathcal{G}) = \mathcal{G}_F = (\mathcal{V} \cup \mathcal{N}, E_F)$ be the corresponding coloured factor graph. Then \mathcal{G} has a supremum in R_V given by

$$s_R(\mathcal{G}) = \phi_V^{-1}((r_{G_F}(\mathcal{V} \cup \mathcal{N}), E_F))$$

We conclude:

Theorem 5.15. $\mathcal{S}_{R_V}^+$ is a complete non-distributive lattice with respect to model inclusion. The meet operation is induced by the meet operation in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9). The join of two models represented by graphs $\mathcal{G}, \mathcal{H} \in R_V$ is represented by the graph $s_R(\mathcal{G} \vee \mathcal{H})$.

Proof: The proof is analogous to the proofs of Theorems 5.3 and 5.6. \square

Lemma 5.12 further implies:

Proposition 5.16. P_V is stable under the meet operation \wedge in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9).

The graphs in Figure 4 also illustrate the stability of P_V under \wedge while establishing that P_V is generally not stable under \vee in $\langle \mathcal{C}_V; \preceq \rangle$. Further:

Lemma 5.17. If $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ and $s_R(\mathcal{G}) = (\mathcal{V}_R, \mathcal{E}_R)$, then for all $\mathcal{G}' = (\mathcal{V}', \mathcal{E}) \in \mathcal{C}_V$ with $\mathcal{V}_R \leq \mathcal{V}' \leq \mathcal{V}$, $s_R(\mathcal{G}') = (\mathcal{V}_R, \mathcal{E}_R)$.

Lemma 5.17 can be shown to imply:

Proposition 5.18. For $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \in \mathcal{C}_V$ let $s_R(\mathcal{G}) = (\mathcal{V}_R, \mathcal{E}_R)$. Then \mathcal{G} has a supremum in P_V given by

$$s_P(\mathcal{G}) = (\mathcal{V}_R, \mathcal{E})$$

We conclude:

Theorem 5.19. $\mathcal{S}_{P_V}^+$ and $\mathcal{R}_{P_V}^+$ are complete lattices with respect to model inclusion. Their meet operation is induced by the meet operation in $\langle \mathcal{C}_V; \preceq \rangle$ given in equation (9). The join of two models represented by $\mathcal{G}, \mathcal{H} \in P_V$ is represented by the graph $s_P(\mathcal{G} \vee \mathcal{H})$.

Proof: In complete analogy to the proofs of Theorems 5.3 and 5.6. \square

6 Model Selection

One way to develop model selection procedures for the model classes considered in this article is by adapting existing model search algorithms for unconstrained graphical Gaussian models. Having shown each of the model classes to be complete lattices, just as the set of standard graphical models, it is natural to consider methods which exploit this structural property. Prominent methods among them are stepwise procedures (Whittaker, 1990; Edwards, 2000), the Edwards–Havránek model selection procedure (Edwards and Havránek, 1987), and, more recently, neighbourhood selection with the lasso (Meinshausen and Bühlmann, 2006), stability selection (Meinshausen and Bühlmann, 2010), and the SINful approach (Drton and Perlman, 2008).

A crucial difference between the search spaces of unconstrained graphical Gaussian models and the models studied here is that while the former constitute a distributive lattice, the latter are all non-distributive. This directly disqualifies neighbourhood selection with the lasso, stability selection and the SINful approach, as they all require distributivity. Put explicitly, while the just mentioned methods are algorithms for determining for each edge whether it is to be present in graph of the accepted model(s) or not, for the model classes considered here not only the edge set needs to be determined, but also partitions of the vertices and present edges into sets corresponding to equal model parameters. This turns model selection into a principally different problem.

For the rest of the article we focus on the Edwards–Havránek model selection procedure and develop a corresponding algorithm for the lattice of models $\mathcal{S}_{B_V}^+$ represented by edge regular colourings. We illustrate the algorithm with a brief summary of its, very encouraging, performance for the data set described in Example 3.1. We further give a summary of an Edwards–Havránek model search within the lattice of models $\mathcal{S}_{\Pi_V}^+$ with permutation-generated colourings for the Fret’s heads data described in Example 4.7. All formal results are given without proof, however they can be obtained by considering the partial ordering of \mathcal{C}_V .

6.1 The Edwards–Havránek Model Selection Procedure

The Edwards–Havránek model selection procedure operates on model search spaces which are lattices and is closely related to the all possible models approach but considerably faster. It is based on the following two principles: (i) if a model is accepted then all models that include it are (weakly) accepted, and (ii) if a model is rejected then all of its submodels are considered to be (weakly) rejected.

The procedure starts by initially testing a set of models and assigns the accepted models to a set \mathcal{A} and the rejected models to set \mathcal{R} . By assumption, all models larger than \mathcal{A} are (weakly) accepted and the ones smaller than \mathcal{R} are (weakly) rejected, so that only $\min \mathcal{A}$, the smallest models in \mathcal{A} , and $\max \mathcal{R}$, the largest models in \mathcal{R} , are of interest. The procedure repeatedly updates $\min \mathcal{A}$ and $\max \mathcal{R}$ and terminates once the set to be updated remains unchanged, when it returns $\min \mathcal{A}$. The method to determine whether a model is to be rejected can be any suitable statistical test in accordance to the principle of *coherence* (Gabriel, 1969), stating that a test should not accept a model while rejecting a larger one.

Following Edwards and Havránek (1987), for a set of models S let $D_a(S)$ denote the set of models in the search lattice L say which are smallest with the property that they are not contained in any model in S ,

$$D_a(S) = \min\{d \in L \mid d \not\subseteq s \text{ for all } s \in S\}$$

and let $D_r(S)$ be the set of largest models that do not contain any model in S ,

$$D_r(S) = \max\{d \in L \mid s \not\subseteq d \text{ for all } s \in S\}$$

$D_a(S)$ is referred to as the *acceptance dual* of S , and $D_r(S)$ as the *rejection dual* of S . The procedure may then be summarised as:

1. Test an initial set of models and assign the accepted models to \mathcal{A} and the rejected models to \mathcal{R} .
2. Choose between 3 and 4.
3. Test the models in $D_r(\mathcal{A}) \setminus \mathcal{R}$. If all are rejected, stop; otherwise, update \mathcal{A} and \mathcal{R} and go to 2.
4. Test the models in $D_a(\mathcal{R}) \setminus \mathcal{A}$. If all are accepted, stop; otherwise, update \mathcal{A} and \mathcal{R} and go to 2.

Acceptance and rejection duals of sets of models can be computed in a recursive manner by using the following two relations. If S and T are two sets of models, then

$$\begin{aligned} D_a(S \cup T) &= \min\{s \vee t \mid s \in D_a(S), t \in D_a(T)\} \\ D_r(S \cup T) &= \max\{s \wedge t \mid s \in D_r(S), t \in D_r(T)\} \end{aligned}$$

Thus describing the duals of a single model is enough.

6.2 Models Represented by Edge Regular Colourings

Proposition 6.1. *Let $\mathcal{S}^+(\mathcal{V}, \mathcal{E}) \in \mathcal{S}_{B_V}^+$ be a model represented by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with edge regular colouring and underlying uncoloured graph $G = (V, E)$. Then the acceptance dual $D_a(\mathcal{S}^+(\mathcal{V}, \mathcal{E}))$ of $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ in $\mathcal{S}_{B_V}^+$ contains all models represented by coloured graphs $\mathcal{G}_a = (\mathcal{V}_a, \mathcal{E}_a)$ satisfying*

- (1i) $\mathcal{V}_a = \{V_1, V_2\}$ such that $\mathcal{V}_a \not\subseteq \mathcal{V}$ and $\mathcal{E}_a = \emptyset$
- (1ii) $\mathcal{V}_a = \{V\}$ and $\mathcal{E}_a = \{E_a\}$ with $\mathcal{E}_a \neq \emptyset$ and $\mathcal{E}_a \not\subseteq \mathcal{E}$.

Put into words, models in the acceptance dual of $\mathcal{S}^+(\mathcal{V}, \mathcal{E}) \in \mathcal{S}_{B_V}^+$ are either represented by the empty graph with two vertex colour classes which are not unions of colour classes in \mathcal{V} , or they are represented by graphs in which all vertices are of the same colour, as are the edges, and the edge set is not a union of colour classes in \mathcal{E} . For example, the coloured graphs in Figure 12 display models which lie in the acceptance dual of the model represented by the edge regular colouring in Figure 5(a).

By definition, acceptance duals are used to test whether there exist models immediately larger than $\max \mathcal{R}$ which can be rejected. Effectively, graphs of type (1i) refine the colouring of the maximally rejected models, while graphs of type (1ii) add edge colour classes, which both give larger models.

Proposition 6.2. *Let $\mathcal{S}^+(\mathcal{V}, \mathcal{E}) \in \mathcal{S}_{B_V}^+$ be a model represented by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with edge regular colouring and underlying uncoloured graph $G = (V, E)$, and for a discrete set A , let $\text{atom}(A)$ denote the partition of A into atomic sets. Then the rejection dual $D_r(\mathcal{S}^+(\mathcal{V}, \mathcal{E}))$ of $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ in $\mathcal{S}_{B_V}^+$ contains all models represented by coloured graphs $\mathcal{G}_r = (\mathcal{V}_r, \mathcal{E}_r)$ satisfying*

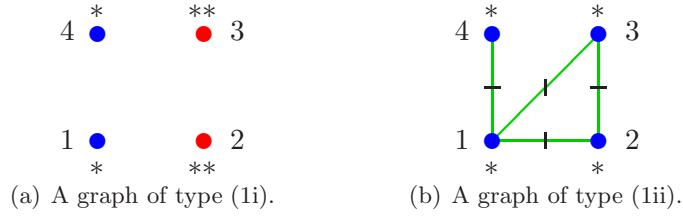


Figure 12: Acceptance dual corresponding to the graph in Figure 5(a).

- (2i) $\mathcal{V}_r = \{\{\alpha, \beta\}\} \cup \text{atom}(V \setminus \{\alpha, \beta\})$ such that $\mathcal{V}_r \not\leq \mathcal{V}$ and $\mathcal{E}_r = \{\alpha\beta \mid \alpha, \beta \in V\}$, or
- (2ii) $\mathcal{V}_r = \text{atom}(V)$ and $\mathcal{E}_r = \text{atom}(\{\alpha\beta \mid \alpha, \beta \in V\} \setminus \{e\})$ with $e \in E$, or
- (2iii) $\mathcal{V}_r = \{\{\alpha, \beta\}, \{\gamma, \delta\}\} \cup \text{atom}(V \setminus \{\alpha, \beta, \gamma, \delta\})$ with α, β and γ, δ being of the same colour in \mathcal{V} and $\mathcal{E}_r = \{\{\alpha\gamma, \beta\delta\}\} \cup \text{atom}(\{\alpha\beta \mid \alpha, \beta \in V\} \setminus \{\alpha\gamma, \beta\delta\})$, where we may have $\alpha = \beta$ or $\gamma = \delta$ but not both, such that $(\mathcal{V}, \mathcal{E}) \not\leq (\mathcal{V}_r, \mathcal{E}_r)$.

Graphs representing models in the rejection dual of a model represented by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ almost represent the unrestricted saturated model except for a minor modification: In graphs of type (2i) two vertices which are not of the same colour in \mathcal{V} form the only composite colour class in \mathcal{V}_r , while graphs of type (2ii) are missing an edge present in $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Graphs of type (2iii) have a pair of equally coloured edges which are not of the same colour in \mathcal{E} , and give the end vertices of the edges an appropriate colouring for the graph colouring to be edge regular. Examples of coloured graphs representing models in the rejection dual of the model represented by the graph displayed in Figure 5(a) are given in Figure 13.

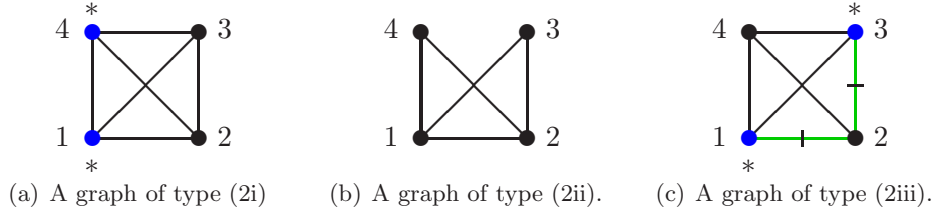


Figure 13: Rejection dual corresponding to the graph in Figure 5(a).

Rejection duals contain models which lie immediately below $\min \mathcal{A}$. Graphs of type (2i) merge vertex colour classes, graphs of type (2ii) cause edge colour classes to be dropped and graphs of type (2iii) merge edge colour classes, as well as vertex colour classes to ensure the resulting model to be edge regular. All operations give graphs which represent smaller models.

It can be shown that while $|D_a(\mathcal{S}^+(\mathcal{V}, \mathcal{E}))|$ grows super-exponentially in the number of variables $|V|$ (at rate $O(2^{|V|^2/2})$), the size of the rejection dual $|D_r(\mathcal{S}^+(\mathcal{V}, \mathcal{E}))|$ grows polynomially in $|V|$ (at rate $O(|V|^4)$), so that from a computational point of view working with rejection duals only is much more efficient. The following algorithm is therefore most efficient for models with edge regular colourings.

1. Test an initial set of models and assign each to \mathcal{R} if it is rejected and to \mathcal{A} otherwise.
2. Test the models in $D_r(\mathcal{A}) \setminus \mathcal{R}$. If all are rejected, stop. Otherwise update \mathcal{A} and \mathcal{R} and repeat.

We executed the above algorithm for the Mathematics marks data set described in Example 3.1 with the saturated uncoloured model as our initial set of accepted models \mathcal{A} and let $\mathcal{R} = \emptyset$ initially. Models were tested for acceptance by performing a likelihood ratio test relative to the saturated unconstrained model at significance level 5% using functionality implemented in the R package `gRc` Højsgaard and Lauritzen (2007). The algorithm fitted 232 models, out of a total of $1.3 \cdot 10^6$, in 8 stages before arriving at 4 minimally accepted models whose graphs are displayed in Figure 14 together with their BIC values. (The 232 models are distributed among the stages as follows. 1: 20 (6 accepted), 2: 21 (19 accepted), 3: 41 (40 accepted), 4: 56 (56 accepted), 5: 55 (55 accepted), 6: 29 (29 accepted), 7: 9 (9 accepted) and 8: 1 (1 accepted).)

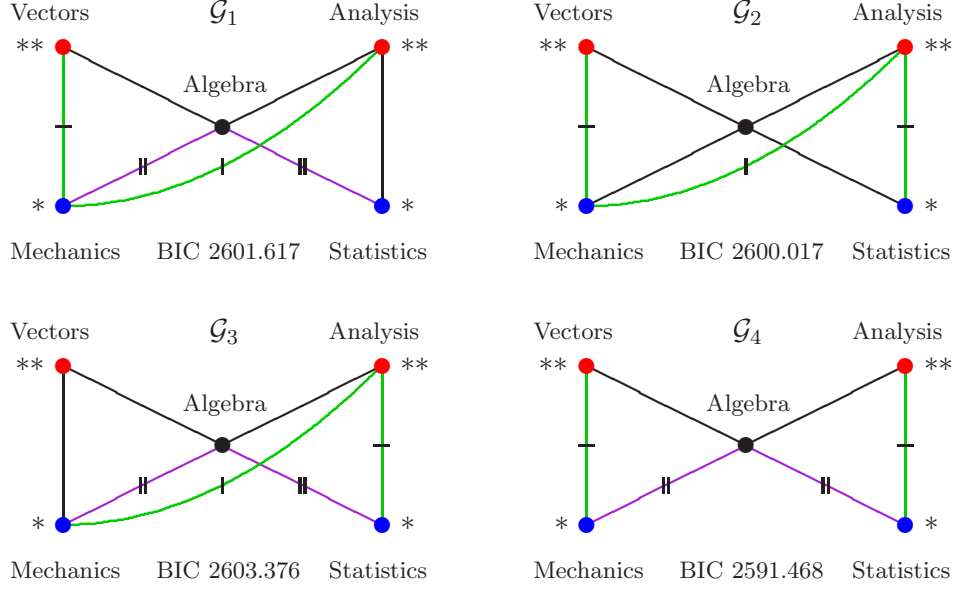


Figure 14: Graphs of minimally accepted models in $\mathcal{S}_{B[5]}^+$.

The uncoloured graphs underlying the graphs in Figure 14 contain the graph in Figure 1(a) as a subgraph. \mathcal{G}_1 , \mathcal{G}_2 and \mathcal{G}_3 only differ by one edge and \mathcal{G}_4 has exactly the same edge set. Thus the conditional independence structures largely agree. The minimally accepted model with the lowest BIC value is represented by \mathcal{G}_4 . It is different from but not dissimilar to the RCON model fitted in Højsgaard and Lauritzen (2008), whose graph is displayed in Figure 1(b), which has a slightly lower BIC value of 2587.404 but no specific properties and it is chosen in an ad hoc manner. Note that the model fitted in Højsgaard and Lauritzen (2008) is not edge regular so that it could not have been considered by the algorithm.

This example suggests that an Edwards–Havráněk model selection procedure for models with edge regular colourings may be feasible in general.

6.3 Models Represented by Permutation-Generated Colourings

The class of permutation-generated colourings Π_V is more complex in its structure than B_V and therefore the duals $D_a(\mathcal{S}^+(\mathcal{V}, \mathcal{E}))$ and $D_r(\mathcal{S}^+(\mathcal{V}, \mathcal{E}))$ of a model $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$ cannot be given in a purely combinatorial form in the graph colouring. A sound understanding of the relationship between a group Γ and its orbits in V and $V \times V$ is required in order to design a general algorithm, in principle, applicable to any variable set V . For illustrative purposes of the

underlying principles we provide a brief summary of an Edwards–Havránek model search in $\mathcal{S}_{\Pi_V}^+$ for Fret’s heads data described in Example 4.7.

For $V = \{B_1, B_2, L_1, L_2\}$ the symmetric group $S(V)$ contains $4! = 24$ permutations and has 30 subgroups, 17 of which are generated by a single permutation. Let \mathcal{K}_V denote the set of colourings of the complete graph these groups generate. The graph colourings in \mathcal{K}_V which are generated by a single permutation, i.e. by $\Gamma = \langle \sigma \rangle$ for $\sigma \in S(V)$, are displayed in Figure 15, where, for the sake of legibility we label the vertices by $V = \{1, 2, 3, 4\}$. The remaining 13 subgroups generate only 5 distinct colourings in $\mathcal{K}_{[4]}$, all of which are shown in Figure 16 with one of their generating groups.

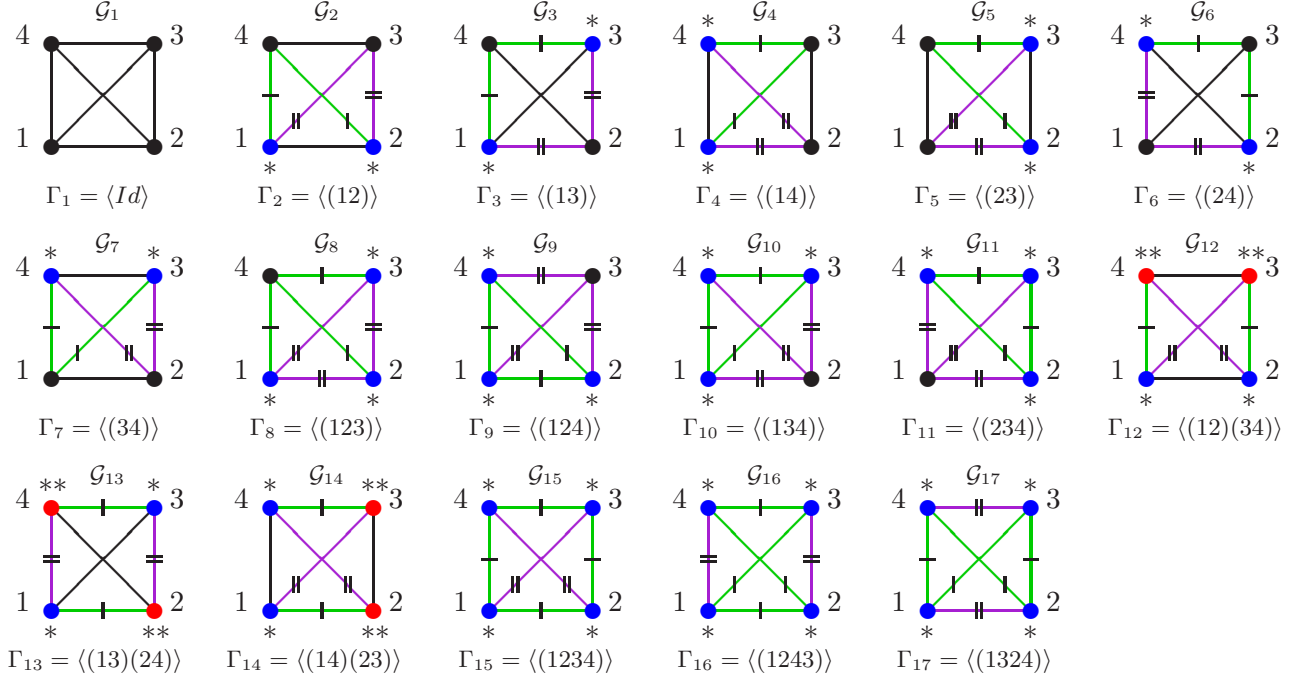


Figure 15: Colourings in $\mathcal{K}_{[4]}$ which are generated by $\Gamma = \langle \sigma \rangle$ for some $\sigma \in S(V)$.

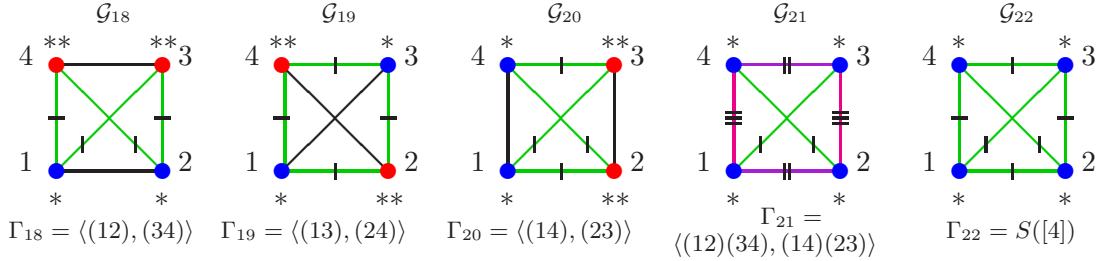


Figure 16: Remaining colourings in $\mathcal{K}_{[4]}$.

The search space $\Pi_{[4]}$ consists of all models which are represented by one of the 22 graphs in $\mathcal{K}_{[4]} = \{\mathcal{G}_1, \dots, \mathcal{G}_{22}\}$ displayed in Figures 15 and 16, together with those represented by graphs which can be obtained from the above by dropping edge colour classes. Thus the size

of the total search space is

$$\begin{aligned}
|\Pi_{[4]}| &= N_1 + 6N_2 + 4N_8 + 3N_{12} + 3(N_{15} - 1) + 3(N_{18} - 4) + (N_{21} - 4) + N_{22} \\
&= 2^6 + 6 \cdot 2^4 + 4 \cdot 2^2 + 3 \cdot 2^4 + 3(2^2 - 1) + 3(2^3 - 4) + (2^3 - 4) + 2 = 251
\end{aligned}$$

where N_i denotes the number of graphs one can obtain from graph \mathcal{G}_i . The subtracted correction terms prevent some graphs to be counted more than once.

Figure 17 displays the Hasse diagram of $\mathcal{K}_{[4]}$. By construction, it contains only graphs which represent the saturated model. An Edwards–Havr nek model selection procedure searches along the full Hasse diagram of all models in $\Pi_{[4]}$, which contains all 251 models and has the diagram in Figure 17 as a subgraph. At each stage, the search moves along the edges in the diagram, passing each model at most once. Once a model has been rejected, all models below it are excluded from the future search; once a model has been accepted all models above it are excluded.

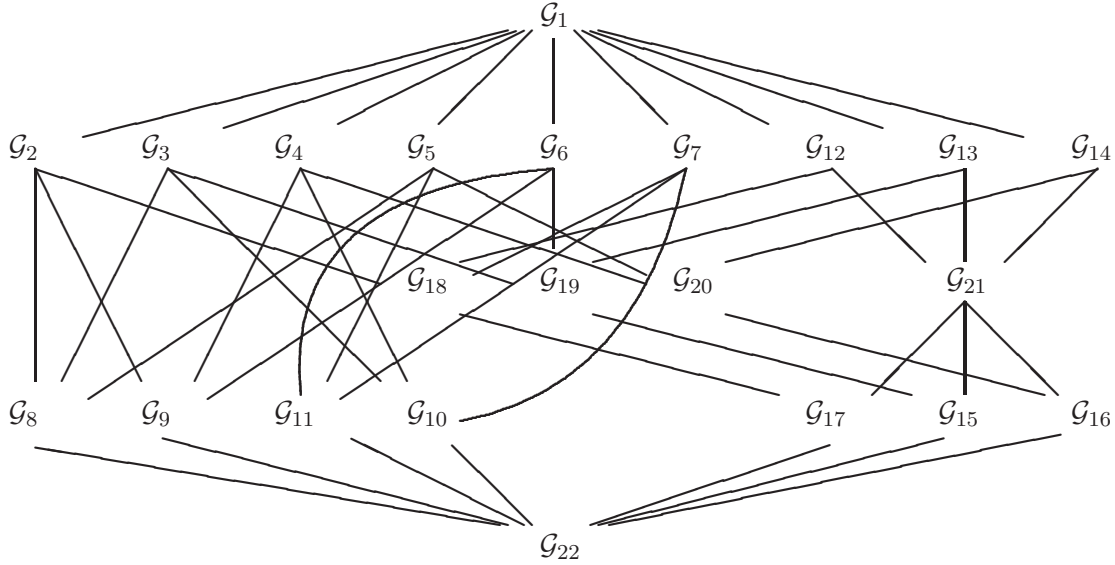


Figure 17: Hasse diagram of $\mathcal{K}_{[4]}$.

Exploiting the demonstrated lattice structure of $\Pi_{[4]}$, we applied the algorithm to the Fret’s heads data described in Example 4.7 with \mathcal{A} initially consisting of the saturated unrestricted model and $\mathcal{R} = \emptyset$. After testing 48 models in 4 stages the algorithm arrived at 9 minimally accepted models whose graphs and generating groups are given in Figure 18. (The models are distributed between the stages the following way. 1: 15 (9 accepted), 2: 16 (16 accepted), 3: 13 (13 accepted), 4: 4 (3 accepted).)

The minimally accepted model with the lowest BIC value is represented by graph \mathcal{H}_9 , which is considerably less than the BIC value 471.2982 of the model fitted in H jsgaard and Lauritzen (2008) whose graph is displayed in Figure 7(b). Further, the model selected in H jsgaard and Lauritzen (2008) is a supermodel, in fact the supremum, of the models represented by \mathcal{H}_7 and \mathcal{H}_8 , with a further edge to complete the four cycle. Interestingly, \mathcal{H}_2 and \mathcal{H}_3 are two of the graphs found in Section 8.3 in Whittaker (1990), the other two being the underlying uncoloured graphs of \mathcal{H}_7 and \mathcal{H}_8 . Note that the BIC value

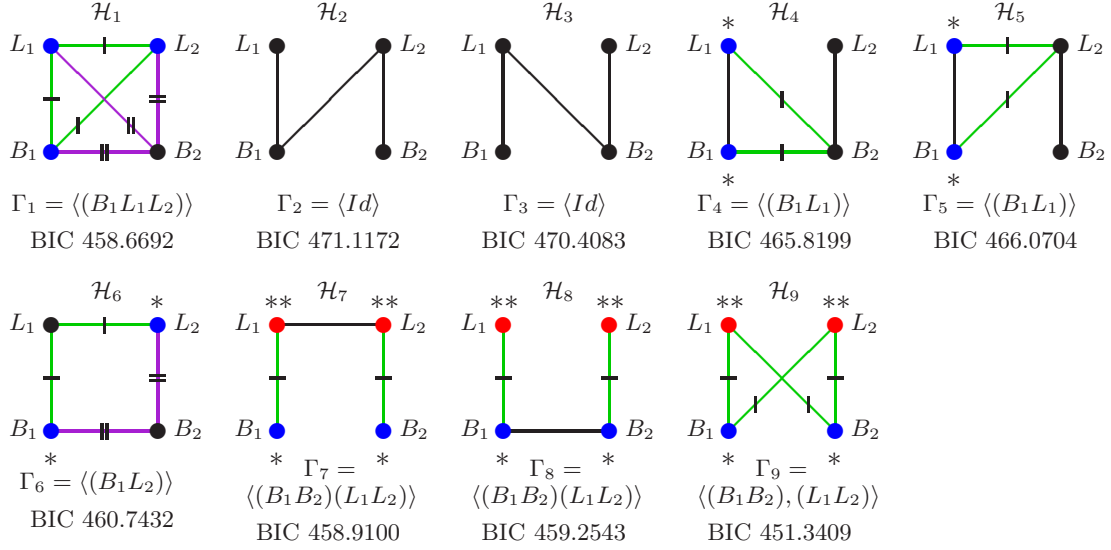


Figure 18: Graphs of minimally accepted models in $\Pi_{[4]}$ for Frets' heads data.

of the complete symmetry model, whose graph is displayed on the right in Figure 7(b), lies between the smallest and the second smallest BIC values of the minimally accepted models, the corresponding graphs being \mathcal{H}_9 and \mathcal{H}_1 respectively. However it was (weakly) rejected by the procedure as it is a submodel of a model rejected in stage 1.

7 Discussion

As we argued, graphical Gaussian models with equality constraints are a promising model class as they combine parsimony in the number of parameters with the concise and efficient graphical models framework. We studied two model types introduced by Højsgaard and Lauritzen (2008): RCON models which place equality restrictions on the model covariance matrix Σ^{-1} and RCOR models which restrict the diagonal of Σ^{-1} and the partial correlations, which can both be represented by vertex and edge coloured graphs.

We showed four model classes within the sets of RCON and RCOR models, each possessing desirable statistical properties and being more readily interpretable than RCON and RCOR models in general, to form complete non-distributive lattices. This qualifies each of them for an Edwards–Havránek model selection procedure. Two model classes, those represented by edge regular and permutation-generated colourings respectively, are most readily interpretable and possess the most tractable structure out of the four and are thus most suitable for a model search.

For the former model class we have developed an Edwards–Havránek model selection algorithm, and demonstrated an encouraging performance for the data set previously described in Example 3.1. We further illustrated the principal functionality of the Edwards–Havránek procedure on the lattice of models represented by permutation-generated colourings with the example of Fret's heads data from Example 4.7. Here as well, the algorithm performed in a satisfactory fashion. In order to fully generalise it to work for any number of variables $|V|$, further investigation into the relationship between permutation groups acting on V and their orbits in V and $V \times V$ is necessary.

Some potential concerns need to be mentioned: firstly, while the algorithm's performance

in the above examples was encouraging, it has to be taken into account that the number of variables is rather small in both cases. Further, it is at this stage unknown how much this behaviour relied on strong/weak conditional independence and symmetry relations in the data sets considered. It may be that the number of models to be tested can still grow in an unmanageable fashion. Secondly, a general concern with the Edwards–Havránek model selection procedure is that its sampling properties are intractable. In particular the procedure does not control the overall error rate.

Especially in view of the above, it may be worthwhile to explore alternative model selection approaches. We argued that neighbourhood selection with the lasso (Meinshausen and Bühlmann, 2006), stability selection (Meinshausen and Bühlmann, 2010), and the SINful approach (Drton and Perlman, 2008) were not directly applicable to the lattices of models studied in this article due to their non-distributivity. Modified variants may still be feasible, which could be investigated.

One further viable alternative may be a symmetry variant of the graphical lasso (Friedman et al., 2008; Ravikumar et al., 2008), which in its original form seeks to maximise the penalised log-likelihood

$$\log \det \Sigma^{-1} - \text{tr}(S\Sigma^{-1}) - \rho \|\Sigma^{-1}\|_1 \quad (12)$$

over non-negative definite matrices Σ^{-1} , with S denoting the empirical covariance matrix of the observations, $\text{tr}(\cdot)$ being the trace, $\|\cdot\|_1$ the l_1 norm giving the sum of the absolute values of the elements in the argument matrix and ρ being the penalisation parameter. Testing for equality constraints on the entries of $\Sigma^{-1} = (k_{\alpha\beta})_{\alpha,\beta \in V}$ can be enforced by replacing equation (12) by the following function

$$\log \det \Sigma^{-1} - \text{tr}(S\Sigma^{-1}) - \rho_1 \|\Sigma^{-1}\|_1 - \rho_2 \sum_{\alpha,\beta \in V} |k_{\alpha\alpha} - k_{\beta\beta}| - \rho_3 \sum_{\substack{\alpha,\beta,\gamma,\delta \in V, \\ \alpha \neq \beta, \gamma \neq \delta}} |k_{\alpha\beta} - k_{\gamma\delta}|$$

This lies in direct analogy to the development of the fused lasso (Tibshirani et al., 2005) from the standard lasso for linear regression. However how to maximise the function for a given model class, for example over models with edge regular colourings to ensure scale invariance, seems non-trivial.

Acknowledgments

I would like to thank my PhD supervisor Steffen Lauritzen for many helpful discussions on the topic and Richard Samworth for inspiring conversations. I am very grateful to Michael Perlman and an anonymous referee for their constructive criticism of an earlier version of this article.

References

Andersen, H., M. Højbjerg, D. Sørensen, and P. Eriksen (1995). *Linear and Graphical Models for the Multivariate Complex Normal Distribution*. Lecture Notes in Statistics. New York, NY, USA: Springer Verlag.

- Anderson, T. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In R. Bose, I. Chakravati, P. Mahalanobis, C. Rao, and K. Smith (Eds.), *Essays in Probability and Statistics*, pp. 1–24. Chapel Hill, NC, USA: University of North Carolina Press.
- Andersson, S. A. (1975). Invariant normal models. *Ann. Statist.* 3, 132–154.
- Bell, E. (1934). Exponential numbers. *American Mathematical Monthly* 1, 411–419.
- Bollobás, B. (1998). *Modern Graph Theory*. New York, NY, USA: Springer Verlag.
- Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Monograph Series. Hayward, CA, USA: Institute of Mathematical Statistics Lecture Notes.
- Buhl, S. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scand. J. Stat.* 20, 263–270.
- Comlet, P. (1974). *Advanced Combinatorics*. Boston, MA, USA: D. Reidel Publishing Company.
- Cox, D. and N. Wermuth (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science* 8, 204–218, 247–277.
- Dobiński, G. (1877). Summierung der Reihe $\sum n^m/n!$ für $m = 1, 2, 3, 4, 5, \dots$. *Grunert Archiv (Archiv für Mathematik und Physik)* 61, 333–336.
- Drton, M. and D. Perlman (2008). A SINful approach to gaussian graphical model selection. *Journal of Statistical Planning and Inference* 138, 1179–1200.
- Edwards, D. (2000). *Introduction to Graphical Modelling*. New York, NY, USA: Springer Verlag.
- Edwards, D. and T. Havránek (1987). A fast model selection procedure for large families of models. *J. Am. Statist. Ass.* 82, 205–213.
- Frets, G. (1921). Heredity of head form in man. *Genetica* 3, 193–400.
- Frey, B. J. (1998). *Graphical Models for Machine Learning and Digital Communication*. Cambridge, MA, USA: The MIT Press.
- Friedman, F., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441.
- Gabriel, K. (1969). Simultaneous test procedures - some theory of multiple comparisons. *Ann. Math. Statist.* 40, 224–250.
- Gehrmann, H. and S. Lauritzen (2011). Estimation of means in graphical Gaussian models with symmetries. <http://arxiv.org/abs/1101.3709>.
- Gottard, A., G. Marchetti, and A. Agresti (2010). Quasi-symmetric graphical log-linear models. *Scand. J. Stat.* no. doi: 10.1111/j.1467-9469.2010.00713.x.

- Grätzer, G. (1998). *General Lattice Theory*. Basel, Switzerland: Birkhäuser Verlag.
- Højsgaard, S. and S. Lauritzen (2007). Inference in graphical Gaussian models with edge and vertex symmetries with the gRc package for R. *Journal of Statistical Software* 23.
- Højsgaard, S. and S. Lauritzen (2008). Graphical Gaussian models with edge and vertex symmetries. *J. R. Statist. Soc. B* 70, 1005–1027.
- Hylleberg, B., M. Jensen, and E. Ørnbøl (1993). Graphical symmetry models. Master’s thesis, Aalborg University, Aalborg.
- Jensen, S. (1988). Covariance hypotheses which are linear in both the covariance and the inverse covariance. *Ann. Statist.* 16, 302–322.
- Lauritzen, S. (1996). *Graphical Models*. Oxford, UK: Clarendon Press.
- Madsen, J. (2000). Invariant normal models with recursive graphical Markov structure. *Ann. Statist.* 28, 1150–1178.
- Mardia, K., J. Kent, and J. Bibby (1979). *Multivariate Analysis*. New York, NY, USA: Academic Press.
- McKay, B. (1976). Backtrack programming and the graph isomorphism problem. Master’s thesis, University of Melbourne.
- Meinshausen, N. and P. Bühlmann (2006). High dimensional graphs and variable selection with the lasso. *Ann. Statist.* 34, 1436–1462.
- Meinshausen, N. and P. Bühlmann (2010). Stability selection. *J. R. Statist. Soc. B* 72, 417–473.
- Olkin, I. (1972). Testing and estimation for structures which are circularly symmetric in blocks. Technical report, Educational Testing Service, Princeton, N.J.
- Olkin, I. and S. J. Press (1969). Testing and estimation for a circular stationary model. *Ann. Math. Statist.* 40, 1358–1373.
- Pitman, J. (1997). Probabilistic aspects of set partitions. *The American Mathematical Monthly* 104, 201–209.
- Ramírez-Aldana, R. (2010). *Restricted or coloured graphical log-linear models*. Ph. D. thesis, Graduate studies in Mathematics, National Autonomous University of Mexico, UNAM.
- Ravikumar, P., M. Wainwright, G. Raskutti, and B. Yu (2008, December). High-dimensional covariance estimation by minimizing l_1 -penalized log-determinant divergence. In *NIPS Conference*, Vancouver, Canada.
- Sachs, H. (1966). Über Teiler, Faktoren und charakteristische Polynome von Graphen. *Wissenschaftliche Zeitschrift der Technischen Hochschule Ilmenau* 12, 7–12.
- Schmidt, R. (1994). *Subgroup Lattices of Groups*. Berlin, Germany: Walter de Gruyter.
- Siemons, J. (1983). Automorphism groups of graphs. *Archiv der Mathematik* 41, 379–384.

- Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight (2005). Sparsity and smoothness via the fused lasso. *J. R. Statist. Soc. B* 67, 91–108.
- Uhler, C. (2010). Geometry of maximum likelihood estimation in Gaussian graphical models. <http://arxiv.org/abs/1012.2643>.
- Votaw, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *Ann. Math. Statist.* 19, 447–473.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester, UK: Wiley.
- Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Math Statist.* 17, 257–281.